Simultaneous Prediction Intervals for Multiple Steps Ahead Forecasts in Vector Time Series

YICK, Kwok Leung

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Master of Philosophy
in
Statistics

July 2007

Thesis/Assessment Committee

Professor CHEUNG Siu Hung (Chair)
Professor WU Ka Ho (Thesis Supervisor)
Professor CHAN Ping Shing (Committee Member)
Professor TSO Kwok Fai Geoffrey (External Examiner)

The Chinese University of Hong Kong
Graduate School

The undersigned certify that we have read a thesis, entitled "Simultaneous Prediction Intervals for Multiple Steps Ahead Forecasts in Vector Time Series" submitted to the Graduate School by Yick Kwok Leung ( ) in partial fulfillment of the requirements for the Degree of Master of Philosophy in Statistics. We recommend that it be accepted.

Prof. K.H. Wu,
Supervisor

Prof. P.S. Chan

Prof. S.H. Cheung

Prof. K.F. Tso,
External Examiner

DECLARATION

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

ACKNOWLEDGMENT

I would like to express my heartfelt gratitude to my supervisor, Prof. Ka-Ho Wu, for his encouragement and invaluable advice during the period of my research program. Moreover, I would like to thank all the staff of the Department of Statistics for their kind helping and supporting. In addition, I wish to acknowledge my fellow classmates and friends of the Department of Statistics for their kind assistance.

ABSTRACT

Forecasting is a vitally important topic in the field of time series analysis, and has been commonly applied in many areas, such as business planning and management. In recent years, multiple forecasting has become more and more popular and is useful for the decision making process. In this thesis, we will focus on how to construct simultaneous prediction intervals for multiple steps ahead forecasts in vector time series. One of the most popular multivariate time series models is the vector autoregressive (VAR) process in which all the variables entering the system have identical order of autoregression. However, we may view the VAR process as a collection of several univariate stationary processes in which each variable may have different number of lags. Based on these two models, we propose two methods for constructing simultaneous prediction intervals for multiple steps ahead forecasts in vector time series. The performances of the various methods will be compared by a simulation study in both two and three-dimensional VAR situations. Two illustrative examples will also be provided.

摘要

在時間序列分析的範疇中，預測是一個非常重要的課題，而它亦被廣泛應用於不同的領域上， 如商業策劃和管理等。近年來，多重的預測越來越流行，並有助於決策。此論文主要集中研究在向量時間序列中 如何構造多重預測的同時預測區間。其中一個最受歡迎的多變數時間序列模型是向量自身迴歸過程， 而所有進入這系統的變數都有相同的自我迴歸序列。但是， 我們可把向量自身迴歸過程看成由多個單變數定常過程組成，而每一個變數都可以有不同的自我迴歸序列。基於這兩個模型， 我們建議兩個方法去構造多重預測的同時預測區間。 我們會利用一個二變數向量自身迴歸模型和一個三變數向量自身迴歸模型來進行模擬研究去比較以上兩個方法的表現。 另外，我們亦會提供兩個説明性的例子。

Contents

Chapter 1

Introduction

1.1 The importance of forecasting

In time series analysis, forecasting has been recognized as an important topic for a long time. It has been commonly applied in many real situations and continously plays an important role in various fields such as finance, marketing and management. For instance, financial institutions have interest in forecasting exchange rates, interest rates and prices of derivatives.

Moreover, forecasting is useful in our daily life. Einhorn and Hogarth (1982) gave the following example of forecasting which is closely related to our daily planning.

' ...Originally, we planned to go for journey to a neighbor province. But the weather forecaster predicted that a typhoon will land on several provinces, including the province we are living in and the one we planned to go within a few days. Therefore, we should cancel our plan and remain at home or perhaps do some precautions, for instance, boarding up property. If it is dangerous to stay at home, it would be better to evacuate to a safety place...'.

In many situations, instead of a single forecast, multiple forecasts are more useful for making decisions. For instance, in order to decide whether it is necessary to reallocate the resources of a company, the company manager may be interested in forecasting all the monthly sales for the coming year simultaneously based on the previous monthly sales records. There is another example of multiple forecasts which can be found in Parigi and Schlitzer (1995). In this article, the data of Italian National Accounts are released by the Italian National Bureau of Statistics with a two-quarters delay. Thus, based on the available early observations, one-step and two-step ahead forecasts of economic activities such as GDP will be helpful for the understanding of the current economy.

Why do we need to construct simultaneous prediction intervals? Alpuim (1997) stated that 'Most of our time, the trajectory of the joint predictions can only be understood by constructing simultaneous prediction intervals.'

In addition, simultaneous prediction intervals are preferred rather than marginal intervals in multiple forecasts. Ravishanker, Wu and Glaz (1991) pointed out that, 'When the problem under investigation is a simultaneous question based on multiple forecasts, the use of marginal prediction intervals will be misleading, since the overall coverage probability is too low.'

1.2 Objective

In recent years, multivariate time series methods for handling a group of related time series have received a lot of attention. For instance, Hang Seng Index, the exchange rate and the interest rate are closely related time series. Thus, multiple time series analysis is needed for understanding the interrelationship among these variables. One of the simplest multivariate time series models is the vector autoregressive (VAR) models in which each variable is explained by its own past values and the past values of all the other variables in the system.

Nevertheless, the vector autoregressive models (VAR) have restricted that all the variables entering the system have identical lag lengths. There is no reason to believe that the same lag length is appropriate for all variables. Hsiao (1981) refined the structure of a VAR system to overcome the mentioned problem in general VAR models. Actually, his original objective was for testing the causality among several time series variables. Based on his idea, we partition the vector autoregressive process into several linear equations. Each equation corresponds to a particular variable in the process and each variable is based on its past records and the past records of other variables, but the lag lengths need not to be the same.

Lütkepohl (1993) and Chan, Cheung and Wu (1999) have discussed how to construct joint forecast regions for vector autoregressive models. In this thesis, our objective is to construct simultaneous prediction intervals for multiple steps ahead forecasts in vector stationary time series. There will be two approaches for achieving the objective above. One is based on the vector autoregressive (VAR) models while another approach is based on a system of linear equations with exogenous variables which does not have any restriction on the lag lengths. In the next chapter, the procedures of constructing simultaneous prediction intervals for the multiple forecasts for VAR models will be described in detail. In Chapter 3, the procedures of constructing simultaneous prediction intervals for the multiple forecasts for a system of linear equation with exogenous variables will be described in detail. Two illustrative examples will be provided in Chapter 4. In Chapter 5, a simulation study is carried out for comparing the performances of the two methods and some concluding remarks are given.

Chapter 2

Vector Autoregressive Model

2.1 The VAR(p) model

Vector autoregressive (VAR) model is widely used for modeling and forecasting multivariate time series data. A k-dimensional VAR model with order p is defined as

$$\mathbf{y}_t = \mathbf{v} + \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t \qquad t = 0,\ \pm 1,\ \pm 2, \ldots \ . \tag{2.1}$$

In equation (2.1), $\mathbf{y}_t = (y_{1,t}, \ldots, y_{k,t})'$ is a random $(k \times 1)$ vector of time series, $\mathbf{v} = (a_{10}, \ldots, a_{k0})'$ is a fixed $(k \times 1)$ vector of intercept terms which allow for the possibility of a nonzero mean $\mathrm{E}(\mathbf{y}_t)$, and $\mathbf{A}_i$ $(i = 1, \ldots, p)$ are fixed $(k \times k)$ coefficient matrices where

$$\mathbf{A}_i = \begin{bmatrix} a_{11,i} & \cdots & a_{1k,i} \\ \vdots & \ddots & \vdots \\ a_{k1,i} & \cdots & a_{kk,i} \end{bmatrix}.$$

The white noise or innovation process is denoted by $\mathbf{u}_t = (u_{1,t}, \ldots, u_{k,t})'$ which is a k-dimensional vector. $\mathbf{u}_t$ is assumed to be independently and identically distributed with

$E(\mathbf{u}_t) = \mathbf{0}$, $E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{\Sigma}_u$ and $E(\mathbf{u}_t\mathbf{u}_s') = \mathbf{0}$ for $t \neq s$. The covariance matrix $\mathbf{\Sigma}_u$ is assumed to be nonsingular. Moreover, $\mathbf{u}_t$ is usually assumed to be Gaussian white noise, i.e. $\mathbf{u}_t \overset{iid}{\sim} N_k(\mathbf{0}, \mathbf{\Sigma}_u)$.

Besides, any VAR(p) model can be written in VAR(1) form. The corresponding kp-dimensional VAR(1) model can be defined as

$$\mathbf{Y}_t = \mathbf{V} + \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{U}_t \tag{2.2}$$

where

$$\mathbf{Y}_t = \begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p+1} \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} \mathbf{v} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{U}_t = \begin{bmatrix} \mathbf{u}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \tag{2.3}$$

are all $(kp \times 1)$ vectors and $\mathbf{A}$ is a $(kp \times kp)$ matrix where

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_k & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_k & \mathbf{0} \end{bmatrix}.$$

The process is stationary if all eigenvalues of matrix $\mathbf{A}$ have modulus less than 1. This condition is equivalent to that the polynomial $\det(\mathbf{I}_{kp} - \mathbf{Az})$ has no root in and on the complex unit circle, or

$$\det(\mathbf{I}_{kp} - \mathbf{Az}) \neq 0 \qquad \text{for } |\mathbf{z}| \leq 1 . \tag{2.4}$$

Under the stability assumption, the VAR(p) model in equation (2.1) has an infinite moving average(MA) representation that is given by

$$\mathbf{y}_t = \mathbf{u} + \sum_{i=0}^{\infty} \mathbf{\Phi}_i \mathbf{u}_{t-i} \tag{2.5}$$

where $\mathbf{u} = (\mathbf{I}_k - \mathbf{A}_1 - \ldots - \mathbf{A}_p)^{-1}\mathbf{v}$ and $\mathbf{\Phi}_i$ (i = 1,2,...) are fixed ($k \times k$) MA coefficient matrices which can be computed recursively by using

$$\mathbf{\Phi}_0 = \mathbf{I}_k \quad \text{and} \quad \mathbf{\Phi}_i = \sum_{j=1}^{i} \mathbf{\Phi}_{i-j}\mathbf{A}_j \tag{2.6}$$

where $\mathbf{A}_j = 0$ for $j > p$ and the elements of $\mathbf{\Phi}_i$ are

$$\mathbf{\Phi}_i = \begin{bmatrix} \phi_{11,i} & \cdots & \phi_{1k,i} \\ \vdots & \ddots & \vdots \\ \phi_{k1,i} & \cdots & \phi_{kk,i} \end{bmatrix}.$$

2.2 Least squares estimation method

In general, there are a number of possible ways for estimating the parameters of VAR(p) models. The multivariate least squares (LS) estimation and the maximum likelihood (ML) estimation are the commonly used methods. Between these two methods, the multivariate least squares (LS) estimation is easier and simpler to use. Moreover, their asymptotic properties (Lütkepohl, 1993, p.66-67, 82-85) are similar. Thus, the multivariate least squares (LS) estimation will be used to estimate the parameters in our study and the maximum likelihood (ML) estimation will not be considered.

Assume that a k-dimensional vector time series $\mathbf{y}_1, \ldots, \mathbf{y}_T$ is available and that is generated by a stationary VAR(p) in equation (2.1). Furthermore, we assume that p pre-sample values $\mathbf{y}_{-p+1}, \ldots, \mathbf{y}_0$ are available.

When the presample values are not available, we simply treat the first p observations as the presample values. Therefore, the actual sample size or the effective series length for a VAR(p) model is $T = n - p$ where n is the original sample size.

We define

$$\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_T) \qquad (k \times T)$$

$$\mathbf{B} = (\mathbf{v}, \mathbf{A}_1, \ldots, \mathbf{A}_p) \qquad (k \times (kp+1))$$

$$Z_t = \begin{bmatrix} 1 \\ \mathbf{y}_t \\ \vdots \\ \mathbf{y}_{t-p+1} \end{bmatrix} \qquad ((kp+1) \times 1) \tag{2.7}$$

$$\mathbf{Z} = (Z_0, \ldots, Z_{T-1}) \qquad ((kp+1) \times T)$$

$$\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_T) \qquad (k \times T).$$

When $\mathbf{y}_1, \ldots, \mathbf{y}_T$ are available, then the VAR(p) model in equation (2.1) can be written as the following form.

$$\mathbf{Y} = \mathbf{B}\mathbf{Z} + \mathbf{U}. \tag{2.8}$$

Nevertheless, the coefficients $\mathbf{v}, \mathbf{A}_1, \ldots, \mathbf{A}_p$ which are included in the matrix $\mathbf{B}$ and the covariance matrix $\mathbf{\Sigma}_u$ are assumed to be unknown. Thus, the time series data will be used to estimate their values. That means $\mathbf{Y}$ and $\mathbf{Z}$ will be used to estimate $\mathbf{B}$ and $\mathbf{\Sigma}_u$. The multivariate least squares (LS) estimator for $\mathbf{B}$ is (see Lütkepohl (1993, p.64))

$$\hat{\mathbf{B}} = (\hat{\mathbf{v}}, \hat{\mathbf{A}}_1, \ldots, \hat{\mathbf{A}}_p) = \mathbf{YZ}'(\mathbf{ZZ}')^{-1} \tag{2.9}$$

where $\mathbf{ZZ}'$ is assumed to be nonsingular.

From section 2.1, we know that $\mathbf{\Sigma}_u = E(\mathbf{u}_t\mathbf{u}_t')$. Thus, a plausible multivariate least squares (LS) estimator for the covariance matrix is (see Lütkepohl (1993, p.67))

$$\tilde{\mathbf{\Sigma}}_u = \frac{1}{T}(\mathbf{Y} - \hat{\mathbf{B}}\mathbf{Z})(\mathbf{Y} - \hat{\mathbf{B}}\mathbf{Z})'. \tag{2.10}$$

Or,

$$\tilde{\mathbf{\Sigma}}_u = \frac{1}{T}(\mathbf{YY}' - \mathbf{YZ}'(\mathbf{ZZ}')^{-1}\mathbf{ZY}'). \tag{2.11}$$

An adjustment for the degrees of freedom is desired due to the fact that in a regression with fixed, nonstochastic regressors, this leads to an unbiased estimator of the covariance matrix. Besides, we can see that there are $kp + 1$ parameters in each of the k equations in equation (2.1). Thus, we obtain the following adjusted estimator for the LS estimated covariance matrix $\tilde{\mathbf{\Sigma}}_u$.

$$\hat{\mathbf{\Sigma}}_u = \frac{T}{T - kp - 1}\tilde{\mathbf{\Sigma}}_u \tag{2.12}$$

The two estimators $\hat{\Sigma}_u$ and $\tilde{\Sigma}_u$ are asymptotically equivalent and their properties can be found in Lütkepohl (1993, p63-69).

2.3 VAR order selection method

In practice, the order p of a VAR model is always unknown and has to be determined. To derive the value of p, we adopt the minimum Akaike's Information Criterion (AIC) which is a commonly used model selection criterion for time series data.

The AIC developed by Akaike (1974) from Fisher's maximum likelihood principle has a great impact to the problems of model identification. Researchers of social and physical sciences always face the difficulty of how to select an adequate order or dimension of a statistical model. The traditional approach for selecting the order is by the hypothesis testing procedures in which a subjective judgment for the decision of levels of significance is required. On the contrary, the AIC method provides researchers an objective and practical way for their problems. In model building, the AIC will decrease when variables are added to the model. At a certain point, the criterion will increase, which is a signal that the added variables may not be necessary. The value of AIC itself is an estimate goodness of fit of the model, so the statistic is attempting to adjust the fit of the model by the number of parameters included. The minimum AIC estimate (MAICE) gives the minimum value of the AIC among all other competing models.

For VAR model, the unknown order is estimated by the MAICE $\hat{p}$ which minimizes the AIC given by

$$\mathrm{AIC}(p) = \ln|\tilde{\mathbf{\Sigma}}_u(p)| + \frac{2pk^2}{T} \tag{2.13}$$

where T is the sample size, k is the dimension of the time series and $\tilde{\mathbf{\Sigma}}_u(p)$ is the least square estimate of $\mathbf{\Sigma}_u$ based on an VAR(p) model for $p = 0, 1, \ldots, P$. The preassigned value P is a finite upper limit of the order of the autoregressive model. If p_0 is the true value of p, it is supposed that $p_0 \leq P < \infty$.

2.4 Constructing simultaneous prediction intervals procedures

Multiple forecasts are useful for the decision making process. For instance, based on previous monthly sales records, a marketing manager is interested in forecasting all monthly sales for the coming year simultaneously. In this section, we are going to discuss how to construct simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1, \ldots, L$ in k-dimensional VAR(p) models and we assume that $\mathbf{y}_t$ follows a Gaussian process and the VAR order p is known. Obviously, the result can be extended to construct simultaneous prediction intervals of $y_{i,t+h}$ where $h = 1, \ldots, L$ for other i.

Assume that the intercept vector $\mathbf{v}$, the coefficient matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$ in (2.1) and the covariance matrix $\mathbf{\Sigma}_u$ are known. Then, the optimal h-step ahead linear minimum mean squared error (MSE) forecast at forecast origin t is (see Lütkepohl (1993, p85))

$$\mathbf{y}_t(h) = \mathbf{v} + \mathbf{A}_1\mathbf{y}_t(h-1) + \ldots + \mathbf{A}_p\mathbf{y}_t(h-p) \tag{2.14}$$

where $\mathbf{y}_t(j) = \mathbf{y}_{t+j}$ for $j \leq 0$ with forecast error being

$$\mathbf{e}_t(h) = \mathbf{y}_{t+h} - \mathbf{y}_t(h) = \sum_{i=0}^{h-1} \mathbf{\Phi}_i \mathbf{u}_{t+h-i} \tag{2.15}$$

where the $\mathbf{\Phi}_i's$ are the coefficient matrices of the MA representation of $\mathbf{y}_t$ defined in (2.6).

The forecast error is distributed as multivariate normal with mean $\mathbf{0}$ and the MSE matrix of $\mathbf{y}_t(h)$ is

$$\mathbf{\Sigma}_y(h) = \sum_{i=0}^{h-1} \mathbf{\Phi}_i \mathbf{\Sigma}_u \mathbf{\Phi}_i'. \tag{2.16}$$

In the following sections, we introduce the Bonferroni procedure and the 'Exact' procedure for constructing the $100(1-\alpha)\%$ simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1, 2, \ldots, L$.

2.4.1 Bonferroni procedure

Lütkepohl (1993, p.34) proposed a conservative method, which is based on the first order Bonferroni inequality, for joint prediction interval construction. Based on his idea, the approximate $100(1-\alpha)\%$ simultaneous prediction intervals for $y_{1,t+h}$ where $h = 1, 2, \ldots, L$ based on Bonferroni procedure is given by

$$(y_{1,t}(h) - z_{(\alpha/2L)}\sigma_{11}(h)\ ,\ y_{1,t}(h) + z_{(\alpha/2L)}\sigma_{11}(h)) \tag{2.17}$$

where $z_{(\alpha/2L)}$ is the $(\alpha/2L)$th upper percentile of the standard normal distribution and $\sigma_{11}(h)$ is the square root of the first diagonal element of $\Sigma_y(h)$ given in (2.16).

2.4.2 The 'Exact' procedure

Let $\mathbf{e}_{1,t,L} = (e_{1,t}(1), \ldots, e_{1,t}(L))'$ follow a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}_{e,L}$ where $e_{1,t}(i)$ is the first element of $\mathbf{e}_t(i)$. Suppose

$$W_{1,t}(h) = e_{1,t}(h)/(Var(e_{1,t}(h)))^{1/2} \qquad \text{for } \ h = 1, \ldots, L \tag{2.18}$$

and let $\mathbf{W}_{1,t,L} = (W_{1,t}(1), \ldots, W_{1,t}(L))'$. Then $\mathbf{W}_{1,t,L}$ follows a multivariate normal distribution with mean $\mathbf{0}$ and correlation matrix $\mathbf{R}_{W,L}$. The element of $\mathbf{R}_{W,L}$ is obtained as in the following.

Consider the m-step ahead forecast error,

$$\mathbf{e}_t(m) = \sum_{i=0}^{m-1} \mathbf{\Phi}_i \mathbf{u}_{t+m-i} \qquad \text{for } m = 1, \ldots, L \tag{2.19}$$

and let $1 \leq m \leq h \leq L$. Then, the h-step ahead forecast error can be rewritten as

$$\begin{aligned} \mathbf{e}_t(h) &= \sum_{i=0}^{h-1} \mathbf{\Phi}_i \mathbf{u}_{t+h-i} \\ &= \sum_{i=0}^{h-m-1} \mathbf{\Phi}_i \mathbf{u}_{t+h-i} + \sum_{i=0}^{m-1} \mathbf{\Phi}_{i+h-m} \mathbf{u}_{t+m-i} \end{aligned} \tag{2.20}$$

and the covariance matrix between $\mathbf{e}_t(h)$ and $\mathbf{e}_t(m)$ is

$$
\begin{aligned}
Cov(\mathbf{e}_t(h), \mathbf{e}_t(m)) &= \sum_{i=0}^{m-1} \mathbf{\Phi}_{i+(h-m)} \mathbf{\Sigma}_u \mathbf{\Phi}_i' \qquad (2.21)\\
&= (s_{ij}(h,m), \quad 1 \leq i,j \leq k).
\end{aligned}
$$

Thus, the covariance between $e_{1,t}(h)$ and $e_{1,t}(m)$ is $s_{11}(h,m)$.

After finding all $s_{11}(h,m)$ for $1 \leq m \leq h \leq L$, we can find out $\mathbf{R}_{W,L}$ which is

$$
\mathbf{R}_{W,L} = (\frac{s_{11}(h,m)}{[s_{11}(h,h)s_{11}(m,m)]^{1/2}} = p_{11}(h,m), \qquad 1 \leq m \leq h \leq L). \qquad (2.22)
$$

The 'Exact' $100(1-\alpha)\%$ simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1, \ldots, L$ are

$$
(y_{1,t}(h) - c\sigma_{11}(h) \ , \ y_{1,t}(h) + c\sigma_{11}(h)) \qquad (2.23)
$$

where c is solved by the following equation

$$
\begin{aligned}
1 - \alpha &= P[|W_{1,t}(h)| \leq c, \ h = 1, \cdots, L]\\
&= \int_{-c}^{c} \cdots \int_{-c}^{c} f(w_{t_1}, \cdots, w_{t_L}) dw_{t_1} \cdots dw_{t_L} \qquad (2.24)
\end{aligned}
$$

where $f(w_{t_1}, \cdots, w_{t_L})$ is the multivariate normal density function of $\mathbf{W}_{1,t,L}$.

In order to construct 'Exact' $100(1-\alpha)\%$ simultaneous prediction intervals, we need to solve equation (2.24) for the value c which depends on the correlation matrix $\mathbf{R}_{W,L}$.

We can use the algorithm by Schervish (1984) to compute the value c. However, it is computationally impractical because it takes extremely long computational time for

$L = 7$ and infeasible for $L \geq 8$ (Glaz and Ravishanker, 1991). The above difficulty can be solved by applying the transformation technique proposed by Genz (1992).

Consequently, the 'Exact' $100(1-\alpha)\%$ simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1, \cdots, L$ can be obtained in equation (2.23).

When the true coefficients of $\mathbf{B} = (\mathbf{v}, \mathbf{A}_1, \ldots, \mathbf{A}_p)$ and $\mathbf{\Sigma}_u$ are unknown, we can simply replace them by $\hat{\mathbf{B}} = (\hat{\mathbf{v}}, \hat{\mathbf{A}}_1, \ldots, \hat{\mathbf{A}}_p)$ and $\hat{\mathbf{\Sigma}}_u$. Moreover, if the true order p is unknown, we may simply replace it by $\hat{p}$ mentioned in section 2.3.

In the following sections, we will write the detailed steps of constructing simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1, 2, 3$ for $k = 2$ and $k = 3$ respectively when the order and all the parameters are unknown.

2.4.3 Two variables case

For $k = 2$, the VAR model with order p can be rewritten as the following equations.

$$
\begin{aligned}
y_{1,t} &= a_{10} + \sum_{l=1}^{p} a_{11,l} y_{1,t-l} + \sum_{l=1}^{p} a_{12,l} y_{2,t-l} + u_{1,t} \\
y_{2,t} &= a_{20} + \sum_{l=1}^{p} a_{21,l} y_{1,t-l} + \sum_{l=1}^{p} a_{22,l} y_{2,t-l} + u_{2,t}
\end{aligned}
$$

where $\mathbf{u} = (u_{1,t}, u_{2,t})'$ is assumed to be Gaussian white noise and we let

$$
\mathbf{\Sigma_u} = \begin{bmatrix} d_1^2 & d_{12}^2 \\ d_{12}^2 & d_2^2 \end{bmatrix}.
$$

Suppose that the order p and all the parameters are unknown and we can simply find out their estimates by following the procedures in section 2.2 and section 2.3. The estimates of all the parameters are

$$\hat{\mathbf{A}}_i = \begin{bmatrix} \hat{a}_{11,i} & \hat{a}_{12,i} \\ \hat{a}_{21,i} & \hat{a}_{22,i} \end{bmatrix} \qquad \text{and} \qquad \hat{\mathbf{v}} = \begin{bmatrix} \hat{a}_{10} \\ \hat{a}_{20} \end{bmatrix}$$

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{u}} = \begin{bmatrix} \hat{d}_1^2 & \hat{d}_{12}^2 \\ \hat{d}_{12}^2 & \hat{d}_2^2 \end{bmatrix}$$

where $i = 1, \ldots, \hat{p}$. The following are the steps of constructing the simultaneous prediction intervals for $y_{1,t+h}$ where $h = 1, 2, 3$.

Step 1: The estimates of the optimal h-step ahead linear minimum mean squared error (MSE) forecasts at forecast origin t for $y_{1,t}$ and $y_{2,t}$ are respectively

$$\hat{y}_{1,t}(h) = \hat{a}_{10} + \sum_{l=1}^{\hat{p}} \hat{a}_{11,l}\hat{y}_{1,t}(h-l) + \sum_{l=1}^{\hat{p}} \hat{a}_{12,l}\hat{y}_{2,t}(h-l)$$
$$\hat{y}_{2,t}(h) = \hat{a}_{20} + \sum_{l=1}^{\hat{p}} \hat{a}_{21,l}\hat{y}_{1,t}(h-l) + \sum_{l=1}^{\hat{p}} \hat{a}_{22,l}\hat{y}_{2,t}(h-l)$$

where $\hat{y}_{i,t}(j) = y_{i,t+j}$ for $j \leq 0$. Note that we need to update both $\hat{y}_{1,t}(h)$ and $\hat{y}_{2,t}(h)$ before updating $\hat{y}_{1,t}(h+1)$ and $\hat{y}_{2,t}(h+1)$. By the above equations, we can calculate the values $\hat{y}_{1,t}(1)$, $\hat{y}_{1,t}(2)$ and $\hat{y}_{1,t}(3)$.

Step 2: By equation (2.16), the estimates of the variances of the 1-step, 2-step and 3-step ahead forecast error can be obtained as follows:

$$
\begin{aligned}
\hat{Var}(e_{1,t}(1)) = \hat{\sigma}_{11}^2(1) &= \hat{d_1^2} \\
\hat{Var}(e_{1,t}(2)) = \hat{\sigma}_{11}^2(2) &= (\hat{a}_{11,1}^2 + 1)\hat{d_1^2} + \hat{a}_{12,1}^2\hat{d_2^2} + 2\hat{a}_{11,1}\hat{a}_{12,1}\hat{d_{12}^2} \\
\hat{Var}(e_{1,t}(3)) = \hat{\sigma}_{11}^2(3) &= ((\hat{a}_{11,1}^2 + \hat{a}_{12,1}\hat{a}_{21,1} + \hat{a}_{11,2})^2 + \hat{a}_{11,1}^2 + 1)\hat{d_1^2} \\
&\quad +((\hat{a}_{11,1}\hat{a}_{12,1} + \hat{a}_{12,1}\hat{a}_{22,1} + \hat{a}_{12,2})^2 + \hat{a}_{12,1}^2)\hat{d_2^2} \\
&\quad +2((\hat{a}_{11,1}^2 + \hat{a}_{12,1}\hat{a}_{21,1} + \hat{a}_{11,2})(\hat{a}_{11,1}\hat{a}_{12,1} + \hat{a}_{12,1}\hat{a}_{22,1} + \hat{a}_{12,2}) \\
&\quad +\hat{a}_{11,1}\hat{a}_{12,1})\hat{d_{12}^2}.
\end{aligned}
$$

Note that $\hat{a}_{ij,l} = 0$ if $l > \hat{p}$.

Step 3: The $100(1-\alpha)\%$ simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1, 2, 3$ based on the Bonferroni procedure can be obtained as follows:

$$(\hat{y}_{1,t}(h) - z_{(\alpha/6)}\hat{\sigma}_{11}(h)\ ,\ \hat{y}_{1,t}(h) + z_{(\alpha/6)}\hat{\sigma}_{11}(h)).$$

Step 4: By using equation (2.21), all the estimates of the covariances can be obtained as follows:

$$
\begin{aligned}
\hat{Cov}(e_{1,t}(2), e_{1,t}(1)) = \hat{s}_{11}(2,1) &= \hat{a}_{11,1}\hat{d_1^2} + \hat{a}_{12,1}\hat{d_{12}^2} \\
\hat{Cov}(e_{1,t}(3), e_{1,t}(1)) = \hat{s}_{11}(3,1) &= (\hat{a}_{11,1}\hat{a}_{11,1} + \hat{a}_{12,1}\hat{a}_{21,1} + \hat{a}_{11,2})\hat{d_1^2} \\
&\quad +(\hat{a}_{11,1}\hat{a}_{12,1} + \hat{a}_{12,1}\hat{a}_{22,1} + \hat{a}_{12,2})\hat{d_{12}^2}
\end{aligned}
$$

$$
\begin{aligned}
\hat{Cov}(e_{1,t}(3), e_{1,t}(2)) = \hat{s}_{11}(3,2) \quad = \quad & (\hat{a}_{11,1}(\hat{a}_{11,1}\hat{a}_{11,1} + \hat{a}_{12,1}\hat{a}_{21,1} + \hat{a}_{11,2}) + \hat{a}_{11,1})\hat{d}_1^2 \\
& + \hat{a}_{12,1}(\hat{a}_{11,1}\hat{a}_{12,1} + \hat{a}_{12,1}\hat{a}_{22,1} + \hat{a}_{12,2})\hat{d}_2^2 \\
& + (\hat{a}_{11,1}(\hat{a}_{11,1}\hat{a}_{12,1} + \hat{a}_{12,1}\hat{a}_{22,1} + \hat{a}_{12,2}) \\
& + \hat{a}_{12,1}(\hat{a}_{11,1}\hat{a}_{11,1} + \hat{a}_{12,1}\hat{a}_{21,1} + \hat{a}_{11,2}) + \hat{a}_{12,1})\hat{d}_{12}^2.
\end{aligned}
$$

Note that $\hat{a}_{ij,l} = 0$ if $l > \hat{p}$. As a result, the estimate $\hat{\mathbf{R}}_{W,3}$ can be found out by putting corresponding estimates in equation (2.22) and note that $\hat{s}_{11}(i,i) = \hat{\sigma}_{11}^2(i)$ for $i = 1, 2, 3$.

Step 5: The 'Exact' $100(1-\alpha)\%$ simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1, 2, 3$ can be obtained as follows:

$$(\hat{y}_{1,t}(h) - c\hat{\sigma}_{11}(h)\ ,\ \hat{y}_{1,t}(h) + c\hat{\sigma}_{11}(h))$$

where c is solved by equation (2.24) and note that $L = 3$.

2.4.4 Three variables case

For $k = 3$, the VAR model with order p can be rewritten as the following equations.

$$
\begin{aligned}
y_{1,t} &= a_{10} + \sum_{l=1}^{p} a_{11,l}y_{1,t-l} + \sum_{l=1}^{p} a_{12,l}y_{2,t-l} + \sum_{l=1}^{p} a_{13,l}y_{3,t-l} + u_{1,t} \\
y_{2,t} &= a_{20} + \sum_{l=1}^{p} a_{21,l}y_{1,t-l} + \sum_{l=1}^{p} a_{22,l}y_{2,t-l} + \sum_{l=1}^{p} a_{23,l}y_{3,t-l} + u_{2,t} \\
y_{3,t} &= a_{30} + \sum_{l=1}^{p} a_{31,l}y_{1,t-l} + \sum_{l=1}^{p} a_{32,l}y_{2,t-l} + \sum_{l=1}^{p} a_{33,l}y_{3,t-l} + u_{3,t}
\end{aligned}
$$

where $\mathbf{u} = (u_{1,t}, u_{2,t}, u_{3,t})'$ is assumed to be Gaussian white noise and we let

$$\mathbf{\Sigma_u} = \begin{bmatrix} d_1^2 & d_{12}^2 & d_{13}^2 \\ d_{12}^2 & d_2^2 & d_{23}^2 \\ d_{13}^2 & d_{23}^2 & d_3^2 \end{bmatrix}.$$

Suppose that the order p and all the parameters are unknown and we can simply find out their estimates by following the procedures in section 2.2 and section 2.3. The estimates of all the parameters are

$$\hat{\mathbf{A}}_i = \begin{bmatrix} \hat{a}_{11,i} & \hat{a}_{12,i} & \hat{a}_{13,i} \\ \hat{a}_{21,i} & \hat{a}_{22,i} & \hat{a}_{23,i} \\ \hat{a}_{31,i} & \hat{a}_{32,i} & \hat{a}_{33,i} \end{bmatrix} \qquad \text{and} \qquad \hat{\mathbf{v}} = \begin{bmatrix} \hat{a}_{10} \\ \hat{a}_{20} \\ \hat{a}_{30} \end{bmatrix}$$

$$\hat{\mathbf{\Sigma}}_\mathbf{u} = \begin{bmatrix} \hat{d}_1^2 & \hat{d}_{12}^2 & \hat{d}_{13}^2 \\ \hat{d}_{12}^2 & \hat{d}_2^2 & \hat{d}_{23}^2 \\ \hat{d}_{13}^2 & \hat{d}_{23}^2 & \hat{d}_3^2 \end{bmatrix}$$

where $i = 1, \ldots, \hat{p}$. The following are the steps of constructing the simultaneous prediction intervals for $y_{1,t+h}$ where $h = 1, 2, 3$.

Step 1: The estimates of the optimal h-step ahead linear minimum mean squared error (MSE) forecast at forecast origin t for $y_{1,t}$, $y_{2,t}$ and $y_{3,t}$ are respectively

$$\hat{y}_{1,t}(h) = \hat{a}_{10} + \sum_{l=1}^{\hat{p}} \hat{a}_{11,l}\hat{y}_{1,t}(h-l) + \sum_{l=1}^{\hat{p}} \hat{a}_{12,l}\hat{y}_{2,t}(h-l) + \sum_{l=1}^{\hat{p}} \hat{a}_{13,l}\hat{y}_{3,t}(h-l)$$
$$\hat{y}_{2,t}(h) = \hat{a}_{20} + \sum_{l=1}^{\hat{p}} \hat{a}_{21,l}\hat{y}_{1,t}(h-l) + \sum_{l=1}^{\hat{p}} \hat{a}_{22,l}\hat{y}_{2,t}(h-l) + \sum_{l=1}^{\hat{p}} \hat{a}_{23,l}\hat{y}_{3,t}(h-l)$$
$$\hat{y}_{3,t}(h) = \hat{a}_{30} + \sum_{l=1}^{\hat{p}} \hat{a}_{31,l}\hat{y}_{1,t}(h-l) + \sum_{l=1}^{\hat{p}} \hat{a}_{32,l}\hat{y}_{2,t}(h-l) + \sum_{l=1}^{\hat{p}} \hat{a}_{33,l}\hat{y}_{3,t}(h-l)$$

where $\hat{y}_{i,t}(j) = y_{i,t+j}$ for $j \leq 0$. Note that we need to update both $\hat{y}_{1,t}(h)$, $\hat{y}_{2,t}(h)$ and $\hat{y}_{3,t}(h)$ before updating $\hat{y}_{1,t}(h+1)$, $\hat{y}_{2,t}(h+1)$ and $\hat{y}_{3,t}(h+1)$. By the above equations, we can calculate the values $\hat{y}_{1,t}(1)$, $\hat{y}_{1,t}(2)$ and $\hat{y}_{1,t}(3)$.

Step 2: By equation (2.16), the estimates of the variances of the 1-step, 2-step and 3-step ahead forecast error can be obtained as follows:

$$
\begin{aligned}
\hat{Var}(e_{1,t}(1)) = \hat{\sigma}_{11}^2(1) &= \hat{d_1^2} \\
\hat{Var}(e_{1,t}(2)) = \hat{\sigma}_{11}^2(2) &= (\hat{a}_{11,1}^2 + 1)\hat{d_1^2} + \hat{a}_{12,1}^2\hat{d_2^2} + \hat{a}_{13,1}^2\hat{d_3^2} + 2\hat{a}_{11,1}\hat{a}_{12,1}\hat{d_{12}^2} \\
&\quad +2\hat{a}_{11,1}\hat{a}_{13,1}\hat{d_{13}^2} + 2\hat{a}_{12,1}\hat{a}_{13,1}\hat{d_{23}^2} \\
\hat{Var}(e_{1,t}(3)) = \hat{\sigma}_{11}^2(3) &= ((\hat{a}_{11,1}\hat{a}_{11,1} + \hat{a}_{12,1}\hat{a}_{21,1} + \hat{a}_{13,1}\hat{a}_{31,1} + \hat{a}_{11,2})^2 + \hat{a}_{11,1}^2 + 1)\hat{d_1^2} \\
&\quad +((\hat{a}_{11,1}\hat{a}_{12,1} + \hat{a}_{12,1}\hat{a}_{22,1} + \hat{a}_{13,1}\hat{a}_{32,1} + \hat{a}_{12,2})^2 + \hat{a}_{12,1}^2)\hat{d_2^2} \\
&\quad +((\hat{a}_{11,1}\hat{a}_{13,1} + \hat{a}_{12,1}\hat{a}_{23,1} + \hat{a}_{13,1}\hat{a}_{33,1} + \hat{a}_{13,2})^2 + \hat{a}_{13,1}^2)\hat{d_3^2} \\
&\quad +2((\hat{a}_{11,1}\hat{a}_{11,1} + \hat{a}_{12,1}\hat{a}_{21,1} + \hat{a}_{13,1}\hat{a}_{31,1} + \hat{a}_{11,2})(\hat{a}_{11,1}\hat{a}_{12,1} \\
&\quad +\hat{a}_{12,1}\hat{a}_{22,1} + \hat{a}_{13,1}\hat{a}_{32,1} + \hat{a}_{12,2}) + \hat{a}_{11,1}\hat{a}_{12,1})\hat{d_{12}^2} \\
&\quad +2((\hat{a}_{11,1}\hat{a}_{12,1} + \hat{a}_{12,1}\hat{a}_{22,1} + \hat{a}_{13,1}\hat{a}_{32,1} + \hat{a}_{12,2})(\hat{a}_{11,1}\hat{a}_{13,1} \\
&\quad +\hat{a}_{12,1}\hat{a}_{23,1} + \hat{a}_{13,1}a_{33,1} + \hat{a}_{13,2}) + \hat{a}_{12,1}\hat{a}_{13,1})\hat{d_{23}^2} \\
&\quad +2((\hat{a}_{11,1}\hat{a}_{11,1} + \hat{a}_{12,1}\hat{a}_{21,1} + \hat{a}_{13,1}\hat{a}_{31,1} + \hat{a}_{11,2})(\hat{a}_{11,1}\hat{a}_{13,1} \\
&\quad +\hat{a}_{12,1}\hat{a}_{23,1} + \hat{a}_{13,1}\hat{a}_{33,1} + \hat{a}_{13,2}) + \hat{a}_{11,1}\hat{a}_{13,1})\hat{d_{13}^2}.
\end{aligned}
$$

Note that $\hat{a}_{ij,l} = 0$ if $l > \hat{p}$.

Step 3: The $100(1-\alpha)\%$ simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1, 2, 3$ based on the Bonferroni procedure can be obtained as follow.

$$(\hat{y}_{1,t}(h) - z_{(\alpha/6)}\hat{\sigma}_{11}(h)\ ,\ \hat{y}_{1,t}(h) + z_{(\alpha/6)}\hat{\sigma}_{11}(h))$$

Step 4: By using equation (2.21), all the estimates of the covariances can be obtained as follows:

$$\begin{aligned}
\hat{Cov}(e_{1,t}(2), e_{1,t}(1)) = \hat{s}_{11}(2,1) &= \hat{a}_{11,1}\hat{d_1^2} + \hat{a}_{12,1}\hat{d_{12}^2} + \hat{a}_{13,1}\hat{d_{13}^2} \\
\hat{Cov}(e_{1,t}(3), e_{1,t}(1)) = \hat{s}_{11}(3,1) &= (\hat{a}_{11,1}\hat{a}_{11,1} + \hat{a}_{12,1}\hat{a}_{21,1} + \hat{a}_{13,1}\hat{a}_{31,1} + \hat{a}_{11,2})\hat{d_1^2} \\
&\quad +(\hat{a}_{11,1}\hat{a}_{12,1} + \hat{a}_{12,1}\hat{a}_{22,1} + \hat{a}_{13,1}\hat{a}_{32,1} + \hat{a}_{12,2})\hat{d_{12}^2} \\
&\quad +(\hat{a}_{11,1}\hat{a}_{13,1} + \hat{a}_{12,1}\hat{a}_{23,1} + \hat{a}_{13,1}\hat{a}_{33,1} + \hat{a}_{13,2})\hat{d_{13}^2} \\
\hat{Cov}(e_{1,t}(3), e_{1,t}(2)) = \hat{s}_{11}(3,2) &= (\hat{a}_{11,1}(\hat{a}_{11,1}\hat{a}_{11,1} + \hat{a}_{12,1}\hat{a}_{21,1} + \hat{a}_{13,1}\hat{a}_{31,1} + \hat{a}_{11,2}) \\
&\quad +\hat{a}_{11,1})\hat{d_1^2} + \hat{a}_{12,1}(\hat{a}_{11,1}\hat{a}_{12,1} + \hat{a}_{12,1}\hat{a}_{22,1} \\
&\quad +\hat{a}_{13,1}\hat{a}_{32,1} + \hat{a}_{12,2})\hat{d_2^2} + \hat{a}_{13,1}(\hat{a}_{11,1}\hat{a}_{13,1} \\
&\quad +\hat{a}_{12,1}\hat{a}_{23,1} + \hat{a}_{13,1}\hat{a}_{33,1} + \hat{a}_{13,2})\hat{d_3^2} \\
&\quad +(\hat{a}_{11,1}(\hat{a}_{11,1}\hat{a}_{12,1} + \hat{a}_{12,1}\hat{a}_{22,1} + \hat{a}_{13,1}\hat{a}_{32,1} \\
&\quad +\hat{a}_{12,2}) + \hat{a}_{12,1}(\hat{a}_{11,1}\hat{a}_{11,1} + \hat{a}_{12,1}\hat{a}_{21,1} \\
&\quad +\hat{a}_{13,1}\hat{a}_{31,1} + \hat{a}_{11,2}) + \hat{a}_{12,1})\hat{d_{12}^2}
\end{aligned}$$

$$+(\hat{a}_{11,1}(\hat{a}_{11,1}\hat{a}_{13,1} + \hat{a}_{12,1}\hat{a}_{23,1} + \hat{a}_{13,1}\hat{a}_{33,1}$$

$$+\hat{a}_{13,2}) + \hat{a}_{13,1}(\hat{a}_{11,1}\hat{a}_{11,1} + \hat{a}_{12,1}\hat{a}_{21,1}$$

$$+\hat{a}_{13,1}\hat{a}_{31,1} + \hat{a}_{11,2}) + \hat{a}_{13,1})\hat{d}_{13}^2$$

$$+(\hat{a}_{12,1}(\hat{a}_{11,1}\hat{a}_{13,1} + \hat{a}_{12,1}\hat{a}_{23,1} + \hat{a}_{13,1}\hat{a}_{33,1}$$

$$+\hat{a}_{13,2}) + \hat{a}_{13,1}(\hat{a}_{11,1}\hat{a}_{12,1} + \hat{a}_{12,1}\hat{a}_{22,1}$$

$$+\hat{a}_{13,1}\hat{a}_{32,1} + \hat{a}_{12,2}))\hat{d}_{23}^2.$$

Note that $\hat{a}_{ij,l} = 0$ if $l > \hat{p}$. As a result, the estimate $\hat{\mathbf{R}}_{W,3}$ can be found by putting corresponding estimates in equation (2.22). Note that $\hat{s}_{11}(i,i) = \hat{\sigma}_{11}^2(i)$ for $i = 1, 2, 3$.

Step 5: The 'Exact' $100(1-\alpha)\%$ simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1, 2, 3$ can be obtained as follows:

$$(\hat{y}_{1,t}(h) - c\hat{\sigma}_{11}(h) \ , \ \hat{y}_{1,t}(h) + c\hat{\sigma}_{11}(h))$$

where c is solved by equation (2.24) and note that $L = 3$.

By using the similar procedures described in this section 2.4, we can also find out the $100(1-\alpha)\%$ simultaneous prediction intervals for the multiple forecasts for other variables.

Chapter 3

A System of Linear Equations with Exogenous Variables

3.1 Restriction of VAR model

The vector autoregressive models (VAR) have restricted that all the variables entering the system have identical orders. This assumption may not generally valid for many time series data especially for economic time series, and may induce bias in the forecasting value. Thus, a system of linear equations is required for allowing each variable with a different number of lags.

Suppose that the vector stationary time series consists of k components, say $\{y_{1,t}, y_{2,t}, \ldots, y_{k,t}\}$. Then their relationship can be written as k sets of linear equations which are

$$y_{1,t} = \beta_{10} + \sum_{l=1}^{m_{11}} \beta_{11,l} y_{1,t-l} + \sum_{l=1}^{m_{12}} \beta_{12,l} y_{2,t-l} + \ldots + \sum_{l=1}^{m_{1k}} \beta_{1k,l} y_{k,t-l} + \varepsilon_{1,t}$$

$$y_{2,t} = \beta_{20} + \sum_{l=1}^{m_{21}} \beta_{21,l} y_{1,t-l} + \sum_{l=1}^{m_{22}} \beta_{22,l} y_{2,t-l} + \ldots + \sum_{l=1}^{m_{2k}} \beta_{2k,l} y_{k,t-l} + \varepsilon_{2,t} \tag{3.1}$$

$$\vdots$$

$$y_{k,t} = \beta_{k0} + \sum_{l=1}^{m_{k1}} \beta_{k1,l} y_{1,t-l} + \sum_{l=1}^{m_{k2}} \beta_{k2,l} y_{2,t-l} + \ldots + \sum_{l=1}^{m_{kk}} \beta_{kk,l} y_{k,t-l} + \varepsilon_{k,t}$$

where m_{ij} are the lag lengths, β_{i0} are constants, $\beta_{ij,l}$ are fixed coefficients and $\{\varepsilon_{1,t}, \varepsilon_{2,t}, \ldots, \varepsilon_{k,t}\}$ are mutually uncorrelated white noise process with finite variances $c_1^2, \ldots, c_k^2$.

Actually, for each equation in equation (3.1), we can treat it as a multiple regression model. For instance, in the first equation in equation (3.1), $y_{1,t-l}, \ldots, y_{k,t-1}$ are the regressors where $y_{1,t}$ is the response. In the following two sections, we will introduce how to estimate the parameters and the lag lengths for each equation in equation (3.1) when the parameters and the lag lengths are unknown.

3.2 Least squares estimation method

In this section, the least squares (LS) estimation applied in multiple regression will be used for estimating the parameters in equations (3.1) and we assume that m_{ij} are known for $1 \leq i, j \leq k$. Moreover, we let $p = max(m_{ij},\ 1 \leq i, j \leq k)$.

Suppose that a k-dimensional vector time series $\mathbf{y}_1, \ldots, \mathbf{y}_T$ is available. Furthermore, we assume that p pre-sample values $\mathbf{y}_{-p+1}, \ldots, \mathbf{y}_0$ are available.

Firstly, we focus on the first linear equation in equation (3.1) which is

$$y_{1,t} = \beta_{10} + \sum_{l=1}^{m_{11}} \beta_{11,l} y_{1,t-l} + \sum_{l=1}^{m_{12}} \beta_{12,l} y_{2,t-l} + \ldots + \sum_{l=1}^{m_{1k}} \beta_{1k,l} y_{k,t-l} + \varepsilon_{1,t} \tag{3.2}$$

where $t = 1, \ldots, T$.

Then, we define

$$\begin{aligned}
\mathbf{Y}_1 &= (y_{1,1}, \ldots, y_{1,T})' && (T \times 1) \\
\mathbf{B}_1 &= (\beta_{10}, \beta_{11,1}, \ldots, \beta_{11,m_{11}}, \ldots, \beta_{1k,1}, \ldots, \beta_{1k,m_{1k}})' && (((m_{11} + \ldots + m_{1k}) + 1) \times 1) \\
\mathbf{U}_1 &= (\varepsilon_{1,1}, \ldots, \varepsilon_{1,T})' && (T \times 1) \qquad (3.3)
\end{aligned}$$

$$Z_{1,t} = \begin{bmatrix} 1 \\ y_{1,t-1} \\ \vdots \\ y_{1,t-m_{11}} \\ \vdots \\ y_{k,t-1} \\ \vdots \\ y_{k,t-m_{1k}} \end{bmatrix} \qquad (((m_{11} + \ldots + m_{1k}) + 1) \times 1)$$

$$\mathbf{Z}_1 = (Z_{1,1}, \ldots, Z_{1,T})' \qquad (T \times ((m_{11} + \ldots + m_{1k}) + 1))$$

The linear equation (3.2) can be written as the following matrix form.

$$\mathbf{Y}_1 = \mathbf{Z_1 B_1} + \mathbf{U_1}. \tag{3.4}$$

The coefficients $\beta_{10}, \beta_{11,1}, \ldots, \beta_{11,m_{11}}, \ldots, \beta_{1k,1}, \ldots, \beta_{1k,m_{1k}}$ which are included in $\mathbf{B_1}$ and the variance of $\varepsilon_{1,t}$ are assumed to be unknown. Thus, the time series data will be used to estimate their values. That means $\mathbf{Y_1}$ and $\mathbf{Z_1}$ will be used to estimate $\mathbf{B_1}$ and c_1^2. The least squares (LS) estimator for $\mathbf{B_1}$ is

$$\hat{\mathbf{B}}_\mathbf{1} = (\hat{\beta}_{10}, \hat{\beta}_{11,1}, \ldots, \hat{\beta}_{11,m_{11}}, \ldots, \hat{\beta}_{1k,1}, \ldots, \hat{\beta}_{1k,m_{1k}})' = (\mathbf{Z_1'Z_1})^{-1}\mathbf{Z_1'Y_1} \tag{3.5}$$

and note that $\mathbf{Z_1'Z_1}$ is assumed to be nonsingular.

Also, the residual sum of squares is

$$SS_{Res} = (\mathbf{Y_1} - \mathbf{Z_1}\hat{\mathbf{B}}_\mathbf{1})'(\mathbf{Y_1} - \mathbf{Z_1}\hat{\mathbf{B}}_\mathbf{1}) \tag{3.6}$$

The unbiased estimator of the variance c_1^2 is given by

$$\hat{c}_1^2 = \frac{1}{T - (m_{11} + \ldots + m_{1k}) - 1}(\mathbf{Y_1} - \mathbf{Z_1}\hat{\mathbf{B}}_\mathbf{1})'(\mathbf{Y_1} - \mathbf{Z_1}\hat{\mathbf{B}}_\mathbf{1}). \tag{3.7}$$

By using the similar procedures in this section, we can estimate all the parameters in the remaining $k-1$ linear equations in equation (3.1).

3.3 Hsiao's sequential method for estimating the lag lengths

In practice, the lag lengths m_{ij} for $1 \le i, j \le k$ are always unknown and have to be determined. Hsiao (1979, 1981) introduced a sequential testing procedure to determine the values of m_{ij} where $1 \le i, j \le k$ for multivariate autoregressive process. Hsiao (1981) applied this idea to U.S. postwar money and income data and found that the model generated by the sequential procedure was useful due to the fact that it can serve as a reduced formulation to avoid imposing false restrictions on the model. Thus, we will use the sequential method to estimate all the lag lengths. In the following sections, we will demostrate the procedures for the two variables case and the three variables case.

3.3.1 Two variables case

For a bivariate autoregressive model, we have the following equations.

$$
\begin{aligned}
y_{1,t} &= \beta_{10} + \sum_{l=1}^{m_{11}} \beta_{11,l} y_{1,t-l} + \sum_{l=1}^{m_{12}} \beta_{12,l} y_{2,t-l} + \varepsilon_{1,t} \\
y_{2,t} &= \beta_{20} + \sum_{l=1}^{m_{21}} \beta_{21,l} y_{1,t-l} + \sum_{l=1}^{m_{22}} \beta_{22,l} y_{2,t-l} + \varepsilon_{2,t}.
\end{aligned}
$$

We set a preassigned value P which is a finite upper limit of all the lag lengths m_{ij}, $i, j = 1, 2$. Suppose that a two-dimensional vector time series $\mathbf{y}_1, \ldots, \mathbf{y}_T$ is available. Furthermore, we assume that P pre-sample values $\mathbf{y}_{-P+1}, \ldots, \mathbf{y}_0$ are available. Without loss of generality, we will only work on $y_{1,t}$ for simplicity. The following steps are the procedures of the sequential method.

Step 1: We consider the equation which regresses $y_{1,t}$ on its own lags.

$$
y_{1,t} = \beta_{10} + \sum_{l=1}^{m_{11}} \beta_{11,l} y_{1,t-l} + \varepsilon_{1,t}.
$$

Step 2: For each $m_{11} = 0, 1, ..., P$, we construct $\mathbf{Y_1}, \mathbf{B_1}$ and $\mathbf{Z_1}$ in equation (3.3) where $k = 2$ and $m_{12} = 0$, and calculate $\hat{\mathbf{B}}_1$ and the residual sum of squares $SS_{Res}(m_{11})$ by equation (3.5) and (3.6) respectively.

Step 3: We adopt the Akaike's (1969) minimum Final prediction error (FPE) as a model selection criterion to determine the optimal lag length. Thus, we calculate

$$\text{FPE}(m_{11}) = \frac{T + m_{11} + 1}{T - m_{11} - 1} \frac{SS_{Res}(m_{11})}{T}$$

for $m_{11} = 0, 1, ..., P$. Then, $\hat{m}_{11}$ is the selected lag length such that $\text{FPE}(\hat{m}_{11}) = \min(\text{FPE}(m_{11}))$ for $m_{11} = 0, 1, ..., P$.

Step 4: We add the other variable and consider the following equation which regresses on the lags of $y_{1,t}$ and $y_{2,t}$ together.

$$y_{1,t} = \beta_{10} + \sum_{l=1}^{\hat{m}_{11}} \beta_{11,l} y_{1,t-l} + \sum_{l=1}^{m_{12}} \beta_{12,l} y_{2,t-l} + \varepsilon_{1,t}.$$

Step 5: For each $m_{12} = 1, ..., P$, we construct $\mathbf{Y_1}, \mathbf{B_1}$ and $\mathbf{Z_1}$ in equation (3.3) and calculate $\hat{\mathbf{B}}_1$ and the residual sum of squares $SS_{Res}(\hat{m}_{11}, m_{12})$ by equation (3.5) and (3.6) respectively.

Step 6: We calculate

$$\text{FPE}(\hat{m}_{11}, m_{12}) = \frac{T + \hat{m}_{11} + m_{12} + 1}{T - \hat{m}_{11} - m_{12} - 1} \frac{SS_{Res}(\hat{m}_{11}, m_{12})}{T}$$

for $m_{12} = 1, ..., P$. Then, $\hat{m}_{12}$ is the selected lag length such that $\text{FPE}(\hat{m}_{11}, \hat{m}_{12}) = \min(\text{FPE}(\hat{m}_{11}, m_{12}))$ for $m_{12} = 1, ..., P$.

Step 7: Finally, we compare $\text{FPE}(\hat{m}_{11})$ and $\text{FPE}(\hat{m}_{11}, \hat{m}_{12})$. If $\text{FPE}(\hat{m}_{11}) < \text{FPE}(\hat{m}_{11}, \hat{m}_{12})$, then the estimated values of m_{11} and m_{12} are $\hat{m}_{11}$ and 0, otherwise the estimated values are $\hat{m}_{11}$ and $\hat{m}_{12}$.

3.3.2 Three variables case

For a trivariate autoregressive model, we have the following equations.

$$\begin{aligned}
y_{1,t} &= \beta_{10} + \sum_{l=1}^{m_{11}} \beta_{11,l} y_{1,t-l} + \sum_{l=1}^{m_{12}} \beta_{12,l} y_{2,t-l} + \sum_{l=1}^{m_{13}} \beta_{13,l} y_{3,t-l} + \varepsilon_{1,t} \\
y_{2,t} &= \beta_{20} + \sum_{l=1}^{m_{21}} \beta_{21,l} y_{1,t-l} + \sum_{l=1}^{m_{22}} \beta_{22,l} y_{2,t-l} + \sum_{l=1}^{m_{23}} \beta_{23,l} y_{3,t-l} + \varepsilon_{2,t} \\
y_{3,t} &= \beta_{30} + \sum_{l=1}^{m_{31}} \beta_{31,l} y_{1,t-l} + \sum_{l=1}^{m_{32}} \beta_{32,l} y_{2,t-l} + \sum_{l=1}^{m_{33}} \beta_{33,l} y_{3,t-l} + \varepsilon_{3,t}.
\end{aligned}$$

Similar to two variables case, we set a preassigned value P which is a finite upper limit of all the lag lengths m_{ij}, $i, j = 1, 2, 3$. Suppose that a three-dimensional vector time series $\mathbf{y}_1, \ldots, \mathbf{y}_T$ is available. Furthermore, we assume that P pre-sample values $\mathbf{y}_{-P+1}, \ldots, \mathbf{y}_0$ are available. Without loss of generality, we will only work on $y_{1,t}$ for simplicity. The following steps are the procedures of the sequential method.

Step 1: Adopt Step 1 to Step 6 in the two variables case.

Step 2: We consider the following equation which regresses on the lags of $y_{1,t}$ and $y_{3,t}$ together.

$$y_{1,t} = \beta_{10} + \sum_{l=1}^{\hat{m}_{11}} \beta_{11,l} y_{1,t-l} + \sum_{l=1}^{m_{13}} \beta_{13,l} y_{3,t-l} + \varepsilon_{1,t}.$$

Step 3: For each $m_{13} = 1, ..., P$, we construct $\mathbf{Y_1}, \mathbf{B_1}$ and $\mathbf{Z_1}$ in equation (3.3) and calculate $\hat{\mathbf{B}}_1$ and the residual sum of squares $SS_{Res}(\hat{m}_{11}, m_{13})$ by equation (3.5) and (3.6) respectively.

Step 4: We calculate

$$\text{FPE}(\hat{m}_{11}, m_{13}) = \frac{T + \hat{m}_{11} + m_{13} + 1}{T - \hat{m}_{11} - m_{13} - 1} \frac{SS_{Res}(\hat{m}_{11}, m_{13})}{T}$$

for $m_{13} = 1, ..., P$. Then, $\hat{m}_{13}$ is the selected lag length such that $\text{FPE}(\hat{m}_{11}, \hat{m}_{13}) = \min(\text{FPE}(\hat{m}_{11}, m_{13}))$ for $m_{13} = 1, ..., P$.

Step 5: Compare the three FPE values which are $\text{FPE}(\hat{m}_{11})$, $\text{FPE}(\hat{m}_{11}, \hat{m}_{12})$ and $\text{FPE}(\hat{m}_{11}, \hat{m}_{13})$.

Case 1) If $\text{FPE}(\hat{m}_{11})$ is the smallest value, then the estimated lag lengths of m_{11}, m_{12} and m_{13} are $\hat{m}_{11}$, 0 and 0 respectively.

Case 2) If $\text{FPE}(\hat{m}_{11}, \hat{m}_{12})$ is the smallest value, then continue the following steps.

1. We add the third variable $y_{3,t}$ and consider the following equation which regresses on the lags of $y_{1,t}$, $y_{2,t}$ and $y_{3,t}$ together.

$$y_{1,t} = \beta_{10} + \sum_{l=1}^{\hat{m}_{11}} \beta_{11,l} y_{1,t-l} + \sum_{l=1}^{\hat{m}_{12}} \beta_{12,l} y_{2,t-l} + \sum_{l=1}^{m_{13}} \beta_{13,l} y_{3,t-l} + \varepsilon_{1,t}$$

2. For each $m_{13} = 1, ..., P$, we construct $\mathbf{Y_1}, \mathbf{B_1}$ and $\mathbf{Z_1}$ in equation (3.3) and calculate $\hat{\mathbf{B}}_1$ and the residual sum of squares $SS_{Res}(\hat{m}_{11}, \hat{m}_{12}, m_{13})$ by equation (3.5) and (3.6) respectively.

3. We calculate

$$\text{FPE}(\hat{m}_{11}, \hat{m}_{12}, m_{13}) = \frac{T + \hat{m}_{11} + \hat{m}_{12} + m_{13} + 1}{T - \hat{m}_{11} - \hat{m}_{12} - m_{13} - 1} \frac{SS_{Res}(\hat{m}_{11}, \hat{m}_{12}, m_{13})}{T}$$

for $m_{13} = 1, ..., P$. Then, $\hat{m}_{13}$ is the selected lag length such that $\text{FPE}(\hat{m}_{11}, \hat{m}_{12}, \hat{m}_{13})$ $= \min(\text{FPE}(\hat{m}_{11}, \hat{m}_{12}, m_{13}))$ for $m_{13} = 1, ..., P$.

4. Compare $\text{FPE}(\hat{m}_{11}, \hat{m}_{12})$ and $\text{FPE}(\hat{m}_{11}, \hat{m}_{12}, \hat{m}_{13})$. If $\text{FPE}(\hat{m}_{11}, \hat{m}_{12}) <$ $\text{FPE}(\hat{m}_{11}, \hat{m}_{12}, \hat{m}_{13})$, then the estimated lag lengths of m_{11}, m_{12} and m_{13} are $\hat{m}_{11}$, $\hat{m}_{12}$ and 0 respectively. Otherwise, the estimated lag lengths of m_{11}, m_{12} and m_{13} are $\hat{m}_{11}$, $\hat{m}_{12}$ and $\hat{m}_{13}$ respectively.

Case 3) If $\text{FPE}(\hat{m}_{11}, \hat{m}_{13})$ is the smallest value, then continue the following steps.

1. We add the third variable $y_{2,t}$ and consider the following equation which regresses on the lags of $y_{1,t}$, $y_{3,t}$ and $y_{2,t}$ together.

$$y_{1,t} = \beta_{10} + \sum_{l=1}^{\hat{m}_{11}} \beta_{11,l} y_{1,t-l} + \sum_{l=1}^{\hat{m}_{13}} \beta_{13,l} y_{3,t-l} + \sum_{l=1}^{m_{12}} \beta_{12,l} y_{2,t-l} + \varepsilon_{1,t}$$

2. For each $m_{12} = 1, ..., P$, we construct $\mathbf{Y_1}, \mathbf{B_1}$ and $\mathbf{Z_1}$ in equation (3.3) and calculate $\hat{\mathbf{B}}_1$ and the residual sum of squares $SS_{Res}(\hat{m}_{11}, \hat{m}_{13}, m_{12})$ by equation (3.5) and (3.6) respectively.

3. We calculate

$$\text{FPE}(\hat{m}_{11}, \hat{m}_{13}, m_{12}) = \frac{T + \hat{m}_{11} + \hat{m}_{13} + m_{12} + 1}{T - \hat{m}_{11} - \hat{m}_{13} - m_{12} - 1} \frac{SS_{Res}(\hat{m}_{11}, \hat{m}_{13}, m_{12})}{T}$$

for $m_{12} = 1, ..., P$. Then, $\hat{m}_{12}$ is the selected lag length such that FPE($\hat{m}_{11}, \hat{m}_{13}, \hat{m}_{12}$) = min(FPE($\hat{m}_{11}, \hat{m}_{13}, m_{12}$)) for $m_{12} = 1, ..., P$.

4. Compare FPE($\hat{m}_{11}, \hat{m}_{13}$) and FPE($\hat{m}_{11}, \hat{m}_{13}, \hat{m}_{12}$). If FPE($\hat{m}_{11}, \hat{m}_{13}$) $<$ FPE($\hat{m}_{11}, \hat{m}_{13}, \hat{m}_{12}$) , then the estimated lag lengths of m_{11}, m_{13} and m_{12} are $\hat{m}_{11}$, $\hat{m}_{13}$ and 0 respectively. Otherwise, the estimated lag lengths of m_{11}, m_{13} and m_{12} are $\hat{m}_{11}$, $\hat{m}_{13}$ and $\hat{m}_{12}$ respectively.

For $y_{2,t}$ and $y_{3,t}$, the procedures can be derived analogously. Furthermore, the hsiao's sequential method can be extended to a general case with k variables by using the similar steps described in this section.

3.4 Using VAR model to construct simultaneous prediction intervals

In this section, we are going to discuss how to construct simultaneous prediction intervals for $y_{1,t+h}$ where $h = 1, \ldots, L$.

In fact, equation (3.1) can be rewritten as a VAR(p) model which is

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{k,t} \end{bmatrix} = \begin{bmatrix} \beta_{10} \\ \beta_{20} \\ \vdots \\ \beta_{k0} \end{bmatrix} + \begin{bmatrix} \beta_{11,1} & \beta_{12,1} & \cdots & \beta_{1k,1} \\ \beta_{21,1} & \beta_{22,1} & \cdots & \beta_{2k,1} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1,1} & \beta_{k2,1} & \cdots & \beta_{kk,1} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{k,t-1} \end{bmatrix} + \cdots$$

$$+\begin{bmatrix} \beta_{11,p} & \beta_{12,p} & \cdots & \beta_{1k,p} \\ \beta_{21,p} & \beta_{22,p} & \cdots & \beta_{2k,p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1,p} & \beta_{k2,p} & \cdots & \beta_{kk,p} \end{bmatrix} \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \vdots \\ y_{k,t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{k,t} \end{bmatrix} \tag{3.8}$$

where $p = max(m_{ij},\ 1 \le i, j \le k)$, $\beta_{ij,l} = 0$ if $l > m_{ij}$ and $\varepsilon_{1,t}, \ldots, \varepsilon_{k,t}$ are assumed to be Gaussian white noises.

Then we let,

$$\mathbf{A}_i = \begin{bmatrix} \beta_{11,i} & \cdots & \beta_{1k,i} \\ \vdots & \ddots & \vdots \\ \beta_{k1,i} & \cdots & \beta_{kk,i} \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} \beta_{10} \\ \vdots \\ \beta_{k0} \end{bmatrix}. \tag{3.9}$$

Then, equation (3.8) can be rewritten as

$$\mathbf{y}_t = \mathbf{v} + \mathbf{A}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t$$

where $\mathbf{y}_t = (y_{1,t}, \ldots, y_{k,t})'$ and $\mathbf{u}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{k,t})'$. Moreover, the covariance matrix of $\mathbf{u}_t$ is

$$\boldsymbol{\Sigma}_{\mathbf{u}} = \begin{bmatrix} c_1^2 & 0 & \cdots & 0 \\ 0 & c_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_k^2 \end{bmatrix}. \tag{3.10}$$

Assume that the intercept vector $\mathbf{v}$, the coefficient matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$ and the covariance matrix $\mathbf{\Sigma}_u$ are known. Then, the optimal h-step ahead linear minimum mean squared error (MSE) forecast at forecast origin t is

$$\mathbf{y}_t(h) = \mathbf{v} + \mathbf{A}_1\mathbf{y}_t(h-1) + \ldots + \mathbf{A}_p\mathbf{y}_t(h-p) \tag{3.11}$$

where $\mathbf{y}_t(j) = \mathbf{y}_{t+j}$ for $j \leq 0$ with forecast error being

$$\mathbf{e}_t(h) = \mathbf{y}_{t+h} - \mathbf{y}_t(h) = \sum_{i=0}^{h-1} \mathbf{\Phi}_i \mathbf{u}_{t+h-i}, \tag{3.12}$$

$\mathbf{\Phi}_i' s$ are the coefficient matrices of the MA representation of $\mathbf{y}_t$ defined in (2.6).

The forecast error is distributed as multivariate normal with mean $\mathbf{0}$ and the MSE matrix of $\mathbf{y}_t(h)$ is

$$\mathbf{\Sigma}_y(h) = \sum_{i=0}^{h-1} \mathbf{\Phi}_i \mathbf{\Sigma}_u \mathbf{\Phi}_i'. \tag{3.13}$$

3.4.1 Bonferroni procedure

Based on section 2.4.1, the conservative simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1, \ldots, L$ based on the Bonferroni procedure are

$$(y_{1,t}(h) - z_{(\alpha/2L)}\sigma_{11}(h)\ ,\ y_{1,t}(h) + z_{(\alpha/2L)}\sigma_{11}(h)) \tag{3.14}$$

where $z_{(\alpha/2L)}$ is the $(\alpha/2L)$th upper percentile of the standard normal distribution.

3.4.2 The 'Exact' procedure

Refer to section 2.4.2, the 'Exact' $100(1-\alpha)\%$ simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1, \ldots, L$ are

$$(y_{1,t}(h) - c(\sigma_{11}(h)) \ , \ y_{1,t}(h) + c(\sigma_{11}(h))) \tag{3.15}$$

where c is solved by equation (2.24) and $\sigma_{11}(h)$ is the square root of the first diagonal element of $\mathbf{\Sigma}_y(h)$. In order to solve equation (2.24) for the value c, we have to compute $\mathbf{R}_{W,L}$. The procedures for finding out $\mathbf{R}_{W,L}$ has been shown in section 2.4.2.

When the $\mathbf{v}$, $\mathbf{A}_1, \ldots, \mathbf{A}_p$ and $\mathbf{\Sigma}_u$ are unknown, we can simply replace them by the estimates which are

$$\hat{\mathbf{A}}_i = \begin{bmatrix} \hat{\beta}_{11,i} & \cdots & \hat{\beta}_{1k,i} \\ \vdots & \ddots & \vdots \\ \hat{\beta}_{k1,i} & \cdots & \hat{\beta}_{kk,i} \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{v}} = \begin{bmatrix} \hat{\beta}_{10} \\ \vdots \\ \hat{\beta}_{k0} \end{bmatrix} \tag{3.16}$$

$$\hat{\mathbf{\Sigma}}_{\mathbf{u}} = \begin{bmatrix} \hat{c}_1^2 & 0 & \cdots & 0 \\ 0 & \hat{c}_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{c}_k^2 \end{bmatrix} \tag{3.17}$$

and note that $\hat{\beta}_{ij,l} = 0$ if $l > m_{ij}$. Furthermore, if the lag lengths m_{ij} for $1 \le i, j \le k$ are unknown, we may simply replace them by $\hat{m}_{ij}$ mentioned in section 3.3. Also, $\hat{p} = max(\hat{m}_{ij},\ 1 \le i, j \le k)$.

In the following sections, we will show the detailed steps of how to construct simultaneous prediction intervals for $y_{1,t+h}$ where $h = 1, 2, 3$ for $k = 2$ and $k = 3$ respectively when all the lag lengths and parameters are unknown.

3.4.3 Two variables case

For $k = 2$, we have the following equations for a bivariate autoregressive model.

$$
\begin{aligned}
y_{1,t} &= \beta_{10} + \sum_{l=1}^{m_{11}} \beta_{11,l} y_{1,t-l} + \sum_{l=1}^{m_{12}} \beta_{12,l} y_{2,t-l} + \varepsilon_{1,t} \\
y_{2,t} &= \beta_{20} + \sum_{l=1}^{m_{21}} \beta_{21,l} y_{1,t-l} + \sum_{l=1}^{m_{22}} \beta_{22,l} y_{2,t-l} + \varepsilon_{2,t}.
\end{aligned}
$$

Suppose that all the lag lengths and all the parameters are unknown and we can simply find out their estimates by following the procedures as in section 3.2 and section 3.3. The following steps are the procedures of constructing the simultaneous prediction intervals for $y_{1,t+h}$ where $h = 1, 2, 3$.

Step 1: The optimal h-step ahead linear minimum mean squared error (MSE) forecast at forecast origin t for $y_{1,t}$ and $y_{2,t}$ are respectively

$$
\begin{aligned}
\hat{y}_{1,t}(h) &= \hat{\beta}_{10} + \sum_{l=1}^{\hat{m}_{11}} \hat{\beta}_{11,l} \hat{y}_{1,t}(h-l) + \sum_{l=1}^{\hat{m}_{12}} \hat{\beta}_{12,l} \hat{y}_{2,t}(h-l) \\
\hat{y}_{2,t}(h) &= \hat{\beta}_{20} + \sum_{l=1}^{\hat{m}_{21}} \hat{\beta}_{21,l} \hat{y}_{1,t}(h-l) + \sum_{l=1}^{\hat{m}_{22}} \hat{\beta}_{22,l} \hat{y}_{2,t}(h-l)
\end{aligned}
$$

where $\hat{y}_{i,t}(j) = y_{i,t+j}$ for $j \leq 0$ and note that we need to update both $\hat{y}_{1,t}(h)$ and $\hat{y}_{2,t}(h)$ before updating $\hat{y}_{1,t}(h+1)$ and $\hat{y}_{2,t}(h+1)$. By using the above equations, we can calculate the values $\hat{y}_{1,t}(1)$, $\hat{y}_{1,t}(2)$ and $\hat{y}_{1,t}(3)$.

Step 2: By equation (3.13), the estimates of variances of the 1-step, 2-step and 3-step ahead forecast error can be obtained.

$$\begin{aligned}
\hat{Var}(e_{1,t}(1)) = \hat{\sigma}_{11}^2(1) &= \hat{c}_1^2 \\
\hat{Var}(e_{1,t}(2)) = \hat{\sigma}_{11}^2(2) &= (\hat{\beta}_{11,1}^2 + 1)\hat{c}_1^2 + \hat{\beta}_{12,1}^2\hat{c}_2^2 \\
\hat{Var}(e_{1,t}(3)) = \hat{\sigma}_{11}^2(3) &= ((\hat{\beta}_{11,1}^2 + \hat{\beta}_{12,1}\hat{\beta}_{21,1} + \hat{\beta}_{11,2})^2 + \hat{\beta}_{11,1}^2 + 1)\hat{c}_1^2 \\
&\quad + ((\hat{\beta}_{11,1}\hat{\beta}_{12,1} + \hat{\beta}_{12,1}\hat{\beta}_{22,1} + \hat{\beta}_{12,2})^2 + \hat{\beta}_{12,1}^2)\hat{c}_2^2
\end{aligned}$$

Note that $\hat{\beta}_{ij,l} = 0$ if $l > \hat{m}_{ij}$.

Step 3: The $100(1-\alpha)\%$ simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1, 2, 3$ based on the Bonferroni procedure can be obtained as follows:

$$(\hat{y}_{1,t}(h) - z_{(\alpha/6)}\hat{\sigma}_{11}(h)\ ,\ \hat{y}_{1,t}(h) + z_{(\alpha/6)}\hat{\sigma}_{11}(h)).$$

Step 4: By using equation (2.21), all the estimates of covariances can be obtained as follows:

$$\begin{aligned}
\hat{Cov}(e_{1,t}(2), e_{1,t}(1)) = \hat{s}_{11}(2,1) &= \hat{\beta}_{11,1}\hat{c}_1^2 \\
\hat{Cov}(e_{1,t}(3), e_{1,t}(1)) = \hat{s}_{11}(3,1) &= (\hat{\beta}_{11,1}\hat{\beta}_{11,1} + \hat{\beta}_{12,1}\hat{\beta}_{21,1} + \hat{\beta}_{11,2})\hat{c}_1^2 \\
\hat{Cov}(e_{1,t}(3), e_{1,t}(2)) = \hat{s}_{11}(3,2) &= (\hat{\beta}_{11,1}(\hat{\beta}_{11,1}\hat{\beta}_{11,1} + \hat{\beta}_{12,1}\hat{\beta}_{21,1} + \hat{\beta}_{11,2}) + \hat{\beta}_{11,1})\hat{c}_1^2 \\
&\quad + \hat{\beta}_{12,1}(\hat{\beta}_{11,1}\hat{\beta}_{12,1} + \hat{\beta}_{12,1}\hat{\beta}_{22,1} + \hat{\beta}_{12,2})\hat{c}_2^2.
\end{aligned}$$

Note that $\hat{\beta}_{ij,l} = 0$ if $l > \hat{m}_{ij}$. As a result, the estimate $\hat{\mathbf{R}}_{W,3}$ can be found by inserting their own estimates in equation (2.22) and note that $\hat{s}_{11}(i,i) = \hat{\sigma}^2_{11}(i)$ for $i = 1, 2, 3$.

Step 5: The 'Exact' $100(1-\alpha)\%$ simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1, 2, 3$ can be obtained as follow

$$(\hat{y}_{1,t}(h) - c\hat{\sigma}_{11}(h)\ ,\ \hat{y}_{1,t}(h) + c\hat{\sigma}_{11}(h))$$

where c is solved by equation (2.24).

3.4.4 Three variables case

For $k = 3$, we have the following equations for a 3-dimensional autoregressive model.

$$\begin{aligned}
y_{1,t} &= \beta_{10} + \sum_{l=1}^{m_{11}} \beta_{11,l} y_{1,t-l} + \sum_{l=1}^{m_{12}} \beta_{12,l} y_{2,t-l} + \sum_{l=1}^{m_{13}} \beta_{13,l} y_{3,t-l} + \varepsilon_{1,t} \\
y_{2,t} &= \beta_{20} + \sum_{l=1}^{m_{21}} \beta_{21,l} y_{1,t-l} + \sum_{l=1}^{m_{22}} \beta_{22,l} y_{2,t-l} + \sum_{l=1}^{m_{23}} \beta_{23,l} y_{3,t-l} + \varepsilon_{2,t} \\
y_{3,t} &= \beta_{30} + \sum_{l=1}^{m_{31}} \beta_{31,l} y_{1,t-l} + \sum_{l=1}^{m_{32}} \beta_{32,l} y_{2,t-l} + \sum_{l=1}^{m_{33}} \beta_{33,l} y_{3,t-l} + \varepsilon_{3,t}.
\end{aligned}$$

Suppose that all the lag lengths and all the parameters are unknown and we can simply find out their estimates by following the procedures as in section 3.2 and section 3.3. The following steps are the procedures of constructing the simultaneous prediction intervals for $y_{1,t+h}$ where $h = 1, 2, 3$.

Step 1: The optimal h-step ahead linear minimum mean squared error (MSE) forecast at forecast origin t for $y_{1,t}$, $y_{2,t}$ and $y_{3,t}$ are respectively

$$\hat{y}_{1,t}(h) = \hat{\beta}_{10} + \sum_{l=1}^{\hat{m}_{11}} \hat{\beta}_{11,l}\hat{y}_{1,t}(h-l) + \sum_{l=1}^{\hat{m}_{12}} \hat{\beta}_{12,l}\hat{y}_{2,t}(h-l) + \sum_{l=1}^{\hat{m}_{13}} \hat{\beta}_{13,l}\hat{y}_{3,t}(h-l)$$

$$\hat{y}_{2,t}(h) = \hat{\beta}_{20} + \sum_{l=1}^{\hat{m}_{21}} \hat{\beta}_{21,l}\hat{y}_{1,t}(h-l) + \sum_{l=1}^{\hat{m}_{22}} \hat{\beta}_{22,l}\hat{y}_{2,t}(h-l) + \sum_{l=1}^{\hat{m}_{23}} \hat{\beta}_{23,l}\hat{y}_{3,t}(h-l)$$

$$\hat{y}_{3,t}(h) = \hat{\beta}_{30} + \sum_{l=1}^{\hat{m}_{31}} \hat{\beta}_{31,l}\hat{y}_{1,t}(h-l) + \sum_{l=1}^{\hat{m}_{32}} \hat{\beta}_{32,l}\hat{y}_{2,t}(h-l) + \sum_{l=1}^{\hat{m}_{33}} \hat{\beta}_{33,l}\hat{y}_{3,t}(h-l)$$

where $\hat{y}_{i,t}(j) = y_{i,t+j}$ for $j \leq 0$ and note that we need to update both $\hat{y}_{1,t}(h)$, $\hat{y}_{2,t}(h)$ and $\hat{y}_{3,t}(h)$ before updating $\hat{y}_{1,t}(h+1)$, $\hat{y}_{2,t}(h+1)$ and $\hat{y}_{3,t}(h+1)$. By using the above equations, we can calculate the values $\hat{y}_{1,t}(1)$, $\hat{y}_{1,t}(2)$ and $\hat{y}_{1,t}(3)$.

Step 2: By equation (3.13), the estimates of the variances of the 1-step, 2-step and 3-step ahead forecast error can be obtained.

$$\begin{aligned}
\hat{Var}(e_{1,t}(1)) = \hat{\sigma}_{11}^2(1) &= \hat{c}_1^2 \\
\hat{Var}(e_{1,t}(2)) = \hat{\sigma}_{11}^2(2) &= (\hat{\beta}_{11,1}^2 + 1)\hat{c}_1^2 + \hat{\beta}_{12,1}^2\hat{c}_2^2 + \hat{\beta}_{13,1}^2\hat{c}_3^2 \\
\hat{Var}(e_{1,t}(3)) = \hat{\sigma}_{11}^2(3) &= ((\hat{\beta}_{11,1}\hat{\beta}_{11,1} + \hat{\beta}_{12,1}\hat{\beta}_{21,1} + \hat{\beta}_{13,1}\hat{\beta}_{31,1} + \hat{\beta}_{11,2})^2 + \hat{\beta}_{11,1}^2 + 1)\hat{c}_1^2 \\
&\quad +((\hat{\beta}_{11,1}\hat{\beta}_{12,1} + \hat{\beta}_{12,1}\hat{\beta}_{22,1} + \hat{\beta}_{13,1}\hat{\beta}_{32,1} + \hat{\beta}_{12,2})^2 + \hat{\beta}_{12,1}^2)\hat{c}_2^2 \\
&\quad +((\hat{\beta}_{11,1}\hat{\beta}_{13,1} + \hat{\beta}_{12,1}\hat{\beta}_{23,1} + \hat{\beta}_{13,1}\hat{\beta}_{33,1} + \hat{\beta}_{13,2})^2 + \hat{\beta}_{13,1}^2)\hat{c}_3^2
\end{aligned}$$

Note that $\hat{\beta}_{ij,l} = 0$ if $l > \hat{m}_{ij}$.

Step 3: The $100(1-\alpha)\%$ simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1,2,3$ based on the Bonferroni procedure can be obtained by equation (3.16) where $L = 3$.

$$(\hat{y}_{1,t}(h) - z_{(\alpha/6)}\hat{\sigma}_{11}(h)\ ,\ \hat{y}_{1,t}(h) + z_{(\alpha/6)}\hat{\sigma}_{11}(h))$$

Step 4: By using equation (2.21), all estimates of the covariances can be obtained as follows:

$$\begin{aligned}
\hat{Cov}(e_{1,t}(2), e_{1,t}(1)) = \hat{s}_{11}(2,1) &= \hat{\beta}_{11,1}\hat{c}_1^2 \\
\hat{Cov}(e_{1,t}(3), e_{1,t}(1)) = \hat{s}_{11}(3,1) &= (\hat{\beta}_{11,1}\hat{\beta}_{11,1} + \hat{\beta}_{12,1}\hat{\beta}_{21,1} + \hat{\beta}_{13,1}\hat{\beta}_{31,1} + \hat{\beta}_{11,2})\hat{c}_1^2 \\
\hat{Cov}(e_{1,t}(3), e_{1,t}(2)) = \hat{s}_{11}(3,2) &= (\hat{\beta}_{11,1}(\hat{\beta}_{11,1}\hat{\beta}_{11,1} + \hat{\beta}_{12,1}\hat{\beta}_{21,1} + \hat{\beta}_{13,1}\hat{\beta}_{31,1} + \hat{\beta}_{11,2}) \\
&\quad +\hat{\beta}_{11,1})\hat{c}_1^2 + \hat{\beta}_{12,1}(\hat{\beta}_{11,1}\hat{\beta}_{12,1} + \hat{\beta}_{12,1}\hat{\beta}_{22,1} \\
&\quad +\hat{\beta}_{13,1}\hat{\beta}_{32,1} + \hat{\beta}_{12,2})\hat{c}_2^2 + \hat{\beta}_{13,1}(\hat{\beta}_{11,1}\hat{\beta}_{13,1} \\
&\quad +\hat{\beta}_{12,1}\hat{\beta}_{23,1} + \hat{\beta}_{13,1}\hat{\beta}_{33,1} + \hat{\beta}_{13,2})\hat{c}_3^2.
\end{aligned}$$

Note that $\hat{\beta}_{ij,l} = 0$ if $l > \hat{m}_{ij}$. As a result, the estimate $\hat{\mathbf{R}}_{W,3}$ can be found out by inserting their own estimates in the equation (2.22) and note that $\hat{s}_{11}(i,i) = \hat{\sigma}_{11}^2(i)$ for $i = 1,2,3$.

Step 5: The 'Exact' $100(1-\alpha)\%$ simultaneous prediction intervals of $y_{1,t+h}$ where $h = 1,2,3$ can be obtained as follows

$$(\hat{y}_{1,t}(h) - c\hat{\sigma}_{11}(h)\ ,\ \hat{y}_{1,t}(h) + c\hat{\sigma}_{11}(h))$$

where c is solved by equation (2.24).

By using the similar procedures described in this section 3.4, we can also find out the $100(1-\alpha)\%$ simultaneous prediction intervals for the multiple forecasts for other variables.

Chapter 4

Illustrative Examples

In this chapter, two illustrative examples will be given. The first example is a two variables case while the second example is a three variables case. We will construct 95% simultaneous predictive intervals for multiple steps ahead forecasts for all the variables in both examples. For simplicity, we use "VAR" to indicate the method based on vector autoregressive model. Also, we use "Linear" to indicate the method based on a system of linear equations with exogenous variables. We will compare the two methods and the results will be displayed in Table 4.1a, 4.1b, 4.2a, 4.2b and 4.2c.

Example 1. We use the sales and a leading indicator in Table B6 of Bowerman and Richard (1987) as our first example and the data are extracted from Box and Jenkins (1976). The data set consists of 150 observations. In order to fit the model with a stationary assumption, we take the first difference of the logarithms of both variables. Therefore, the data set consists of 149 observations after differencing. We let "Leading indicator" be the first variable $y_{1,t}$ while "Sales" be the second variable $y_{2,t}$ where $t = 1, ..., 149$. The upper bound of the order and all the lag lengths are set at $P = 5$. Thus, the first five values are set as the presample values. Moreover, in order to evaluate the performance of the the methods mentioned before, the last 3 observations $y_{1,146+h}$ and $y_{2,146+h}$ where $h = 1, 2, 3$ are assigned to the 1-step, 2-step and 3-step ahead forecast period and considered as the holdout period. As a result, the sample size is $T = 141$. We will construct 95% simultaneous prediction intervals for $y_{1,146+h}$ and $y_{2,146+h}$ respectively where $h = 1, 2, 3$ by using the methods mentiond before.

Firstly, we focus on method based on vector autoregressive model. We assume that the data is modelled by VAR process and the order p and all the parameters of the VAR model are unknown. Thus, we need to make use of section 2.2 and 2.3 to estimate all the parameters and the order p. We find out that

$$\hat{p} = 5.$$

After estimating the order and all the parameters, we follow the steps shown in section 2.4.3 for constructing the 95% simultaneous prediction intervals for $y_{1,146+h}$ where $h = 1, 2, 3$ based on Bonferroni and Exact procedures respectively. Then, we repeat the steps analogously for the second variable.

Secondly, we focus on the method based on a system of linear equations with exogenous variables. We assume that the data is modelled by a system of linear equations with exogenous variables and all the lag lengths and parameters are unknown. By following the procedures in section 3.2 and 3.3, we can find out all the estimates of the lag lengths and the parameters. The estimated lag lengths are in the following.

$$\hat{m}_{11} = 3 \qquad \text{and} \qquad \hat{m}_{12} = 0$$

$$\hat{m}_{21} = 5 \qquad \text{and} \qquad \hat{m}_{22} = 4.$$

After estimating all the parameters and the lag lengths, we follow the steps shown in section 3.4.3 for constructing the 95% simultaneous prediction intervals for $y_{1,146+h}$ where $h = 1, 2, 3$ based on Bonferroni and Exact procedures respectively. Then, we repeat the steps analogously for the second variable.

The normality assumption in this example has already been checked to be valid.

Table 4.1a provides 95% simultaneous prediction intervals for the first variable "Leading indicator" and the actual observations $y_{1,146+h}$ where $h = 1, 2, 3$ based on both "VAR" and "Linear" methods and the two procedures Bonferroni and Exact. Table 4.1b shows out the 95% simultaneous predicion intervals for the second variable "Sales" and the actual observations $y_{2,146+h}$ where $h = 1, 2, 3$ based on both "VAR" and "Linear" methods and the two procedures Bonferroni and Exact.

From Table 4.1a, all the four types of 95% simultaneous prediction intervals cover the actual observations $y_{1,146+h}$ where $h = 1, 2, 3$. However, all the widths of the 95% simultaneous prediction intervals of "Linear" method are narrower than the widths of the prediction intervals of "VAR" method under both the Bonferroni and Exact procedure respectively.

From Table 4.1b, all the four types of 95% simultaneous prediction intervals cover the actual observations $y_{2,146+h}$ where $h = 1, 2, 3$. Also, all the widths of the 95% simultaneous prediction intervals of "Linear" method are the same as the widths of the prediction intervals of "VAR" method under both Bonferroni and Exact procedure respectively.

In this example, "Linear" method is better than the "VAR" method for constructing simultaneous prediction intervals because it can provide narrower widths of the intervals for the first variable.

Example 2. We adopt the west German fixed investment, disposable income and consumption expenditures in Table E.1 of Lütkepohl (1993) as our second example. We use data from 1960 to 1982 and there are totally 92 observations. As we want to fit a model with stationary assumption, we take the first difference of the logarithms of all variables. We set "Investment" as the first variable $y_{1,t}$ where $t = 1, ..., 91$ while "Income" and "Consumption" are respectively the second variable $y_{2,t}$ and third variable $y_{3,t}$ where $t = 1, ..., 91$. Also, the upper bound of all the orders are set at $P = 5$ and thus the first five values are set as the presample values. Furthermore, in order to evaluate the performances of the methods mentioned before, the last 3 observations $y_{1,88+h}$, $y_{2,88+h}$ and $y_{3,88+h}$ where $h = 1, 2, 3$ are assigned to the 1-step, 2-step and 3-step ahead forecast period as $t = 89, 90, 91$ are considered as the holdout period. As a result, the sample size is $T = 83$. We will construct 95% simultaneous prediction intervals for $y_{1,88+h}$, $y_{2,88+h}$ and $y_{3,88+h}$ respectively where $h = 1, 2, 3$ by using the methods mentiond before.

Firstly, we focus on the method based on vector autoregressive model. We assume that the data is modelled by VAR process and the order p and all the parameters of the VAR model are unknown. Thus, we need to make use of section 2.2 and 2.3 to estimate all the parameters and the order p. We find out that

$$\hat{p} = 2.$$

After estimating the order and all the parameters, we follow the steps shown in section 2.4.4 for constructing the 95% simultaneous prediction intervals for $y_{1,88+h}$ where $h = 1, 2, 3$ based on Bonferroni and Exact procedure respectively. Then, we repeat the steps analogously for the the second variable and the third variable.

Secondly, we focus on method based on a system of linear equations with exogenous variables. We assume that the data is modelled by a system of linear equations with exogenous variables and all the lag lengths and parameters are unknown. By following the procedures in section 3.2 and 3.3, we can find out all the estimates of the lag lengths and the parameters. The estimated lag lengths are as in the following:

$$\hat{m}_{11} = 4 \quad , \quad \hat{m}_{12} = 1 \quad , \quad \hat{m}_{13} = 0$$

$$\hat{m}_{21} = 1 \quad , \quad \hat{m}_{22} = 3 \quad , \quad \hat{m}_{23} = 0$$

$$\hat{m}_{31} = 0 \quad , \quad \hat{m}_{32} = 3 \quad , \quad \hat{m}_{33} = 3.$$

After estimating all the parameters and lag lengths, we follow the steps shown in section 3.4.4 for constructing the 95% simultaneous prediction intervals for $y_{1,88+h}$ where $h = 1, 2, 3$ based on Bonferroni and Exact procedure respectively. Then, we repeat the steps analogously for the second variable and the third variable.

The normality assumption in this example has already been checked to be valid.

Table 4.2a shows out all the 95% simultaneous prediction intervals for the first variable "Investment" while Table 4.2b and 4.2c show out all the 95% simultaneous predicion intervals for the second variable "Income" and the third variable "Consumption" respectively.

For each of the three variables, all the four types of 95% simultaneous prediction intervals cover the actual observations. However, all the widths of the 95% simultaneous prediction intervals of "Linear" method are narrower than the widths of the prediction intervals of "VAR" method under the Bonferroni and Exact procedure respectively.

In this example, "Linear" method is better than the "VAR" method because it can provide narrower widths of simultaneous prediction intervals for all the three variables.

In the next chapter, we will evaluate their performances by a simulation study.

Table 4.1a 95% simultaneous prediction intervals for the first variable in Example 1

Method	Procedure	h-step ahead forecast	Lower bound	Upper bound	Interval width	Actual Observation
VAR	Bonferroni	1	-0.0627	0.0557	0.1184	-0.0052
		2	-0.0646	0.0668	0.1313	0.0191
		3	-0.0602	0.0716	0.1318	-0.0272
	Exact	1	-0.0621	0.0551	0.1172	-0.0052
		2	-0.0639	0.0661	0.1300	0.0191
		3	-0.0595	0.0709	0.1304	-0.0272
Linear	Bonferroni	1	-0.0599	0.0567	0.1166	-0.0052
		2	-0.0640	0.0660	0.1299	0.0191
		3	-0.0622	0.0678	0.1300	-0.0272
	Exact	1	-0.0593	0.0561	0.1153	-0.0052
		2	-0.0633	0.0653	0.1286	0.0191
		3	-0.0615	0.0671	0.1287	-0.0272

Table 4.1b 95% simultaneous prediction intervals for the second variable in Example 1

Method	Procedure	h-step ahead forecast	Lower bound	Upper bound	Interval width	Actual Observation
VAR	Bonferroni	1	-0.0070	-0.0015	0.0055	-0.0038
		2	-0.0006	0.0050	0.0056	0.0015
		3	0.0000	0.0058	0.0058	0.0019
	Exact	1	-0.0070	-0.0015	0.0055	-0.0038
		2	-0.0006	0.0050	0.0056	0.0015
		3	0.0000	0.0057	0.0058	0.0019
Linear	Bonferroni	1	-0.0072	-0.0017	0.0055	-0.0038
		2	-0.0008	0.0048	0.0056	0.0015
		3	-0.0003	0.0055	0.0058	0.0019
	Exact	1	-0.0071	-0.0017	0.0055	-0.0038
		2	-0.0007	0.0048	0.0056	0.0015
		3	-0.0003	0.0055	0.0058	0.0019

Table 4.2a 95% simultaneous prediction intervals for the first variable in Example 2

Method	Procedure	h-step ahead forecast	Lower bound	Upper bound	Interval width	Actual Observation
VAR	Bonferroni	1	-0.0911	0.1234	0.2145	0.0283
		2	-0.1040	0.1194	0.2233	0.0085
		3	-0.1007	0.1236	0.2243	-0.0012
	Exact	1	-0.0908	0.1230	0.2138	0.0283
		2	-0.1036	0.1190	0.2225	0.0085
		3	-0.1003	0.1232	0.2235	-0.0012
Linear	Bonferroni	1	-0.0760	0.1291	0.2051	0.0283
		2	-0.1041	0.1089	0.2130	0.0085
		3	-0.1167	0.0965	0.2132	-0.0012
	Exact	1	-0.0756	0.1286	0.2042	0.0283
		2	-0.1036	0.1084	0.2120	0.0085
		3	-0.1162	0.0960	0.2122	-0.0012

Table 4.2b 95% simultaneous prediction intervals for the second variable in Example 2

Method	Procedure	h-step ahead forecast	Lower bound	Upper bound	Interval width	Actual Observation
VAR	Bonferroni	1	-0.0154	0.0388	0.0542	-0.0080
		2	-0.0113	0.0449	0.0563	0.0038
		3	-0.0107	0.0464	0.0571	0.0087
	Exact	1	-0.0153	0.0388	0.0541	-0.0080
		2	-0.0113	0.0449	0.0561	0.0038
		3	-0.0107	0.0463	0.0569	0.0087
Linear	Bonferroni	1	-0.0112	0.0425	0.0538	-0.0080
		2	-0.0090	0.0454	0.0544	0.0038
		3	-0.0117	0.0429	0.0546	0.0087
	Exact	1	-0.0112	0.0424	0.0536	-0.0080
		2	-0.0089	0.0453	0.0543	0.0038
		3	-0.0116	0.0429	0.0545	0.0087

Table 4.2c 95% simultaneous prediction intervals for the third variable in Example 2

Method	Procedure	h-step ahead forecast	Lower bound	Upper bound	Interval width	Actual Observation
VAR	Bonferroni	1	-0.0074	0.0396	0.0470	0.0009
		2	-0.0129	0.0379	0.0508	0.0058
		3	-0.0088	0.0443	0.0531	0.0093
	Exact	1	-0.0073	0.0395	0.0468	0.0009
		2	-0.0128	0.0378	0.0507	0.0058
		3	-0.0087	0.0442	0.0529	0.0093
Linear	Bonferroni	1	-0.0031	0.0414	0.0444	0.0009
		2	-0.0136	0.0371	0.0506	0.0058
		3	-0.0114	0.0414	0.0527	0.0093
	Exact	1	-0.0029	0.0412	0.0441	0.0009
		2	-0.0133	0.0368	0.0502	0.0058
		3	-0.0112	0.0411	0.0523	0.0093

Chapter 5

A Simulation Study

In this chapter, we try to evaluate the performance of the methods described in the previous chapters. We provide a simulation study and compare the results of all the methods when constructing 95% simultaneous prediction intervals for the multiple steps ahead forecasts in two and three-dimensional VAR situations. Similar to chapter 4, we use "VAR" to represent the method based on vector autoregressive model and we use "Linear" to represent the method based on a system of linear equations with exogenous variables.

5.1 Design of the experiment

The following two models are chosen for the simulation.

Model 1 : A two-dimensional VAR(2) process

$$\mathbf{y}_t = \begin{pmatrix} 0.0 \\ 0.0 \end{pmatrix} + \begin{pmatrix} 0.50 & 0.00 \\ 0.20 & 0.40 \end{pmatrix} \mathbf{y}_{t-1} + \begin{pmatrix} 0.30 & 0.00 \\ 0.20 & 0.00 \end{pmatrix} \mathbf{y}_{t-2} + \mathbf{u}_t$$

$$\text{with } \boldsymbol{\Sigma}_u = \begin{pmatrix} 0.90 & 0.10 \\ 0.10 & 0.80 \end{pmatrix}.$$

Model 2 : A three-dimensional VAR(2) process

$$\mathbf{y}_t = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \end{pmatrix} + \begin{pmatrix} 0.40 & 0.00 & 0.00 \\ 0.20 & 0.40 & 0.00 \\ 0.00 & 0.30 & 0.40 \end{pmatrix} \mathbf{y}_{t-1} + \begin{pmatrix} 0.30 & 0.00 & 0.00 \\ 0.20 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.30 \end{pmatrix} \mathbf{y}_{t-2} + \mathbf{u}_t$$

$$\text{with } \boldsymbol{\Sigma}_u = \begin{pmatrix} 1.20 & 0.10 & 0.20 \\ 0.10 & 1.00 & 0.00 \\ 0.20 & 0.00 & 1.00 \end{pmatrix}.$$

In each model, $\mathbf{u}_t$'s are identically, idependently and normally distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_u$. Now, we first outline the steps for the simulation study of Model 1.

Step 1: Input $\mathbf{v}$, $\mathbf{A}_i (i = 1, \ldots, p)$, $\boldsymbol{\Sigma}_u$ of Model 1.

Step 2: 353 observations $\mathbf{y}_t$, $t = 1, \ldots, 353$ are generated from the model.

Step 3: Discard the first 200 observations $\mathbf{y}_t$ to reduce the impact of starting-up values, and the last 3 observations are assigned to the 1-step, 2-step and 3-step ahead forecast period for evaluating the performance of forecasts. Thus, n is equal to 150.

Step 4: The maximum order and lag length P is set to 5, so the first 5 observations are set aside for the presample values. Therefore $T = 150 - 5 = 145$.

The "VAR" method and the "Linear" method will be applied to the generated data respectively. Each of their steps are listed below.

For "VAR" method:

Step 5a: Construct $\mathbf{Y}$ and $\mathbf{Z}$ by equation (2.7).

Step 6a: Compute $\hat{\mathbf{B}} = (\hat{\mathbf{v}}, \hat{\mathbf{A}}_1, \ldots, \hat{\mathbf{A}}_p)$ by equation (2.9).

Step 7a: Evaluate $\tilde{\mathbf{\Sigma}}_u$ by equation (2.10) and adjust it by equation (2.12).

Step 8a: Compute the AIC value by equation (2.13).

Step 9a: Repeat Step 5a to Step 8a for $p = 0, 1, 2, 3, 4, 5$.

Step 10a: Choose the order $\hat{p}$ which minimize the AIC.

Step 11a: Adopt the steps in section 2.4.3.

Step 12a: Construct the 95% simultaneous prediction intervals based on the Bonferroni procedure of $y_{1,350+h}$ where $h = 1, 2, 3$. The width of the h-step ahead forecast is calculated as

$$2z_{(0.05/6)}\hat{\sigma}_{11}(h)$$

for $h = 1, 2, 3$ where $z_{(0.05/6)}$ is the $(0.05/6)$th upper percentile of the standard normal distribution.

Step 13a: Construct the 95% simultaneous prediction intervals based on the Exact procedure of $y_{1,350+h}$ where $h = 1, 2, 3$. The width of the h-step ahead forecast is calculated as

$$2c\hat{\sigma}_{11}(h)$$

for $h = 1, 2, 3$.

Step 14a: Repeat Step 11a to 13a analogously for $y_{2,t}$.

Step 15a: Repeat Step 2 to Step 4 and Step 5a to Step 14a for NR replications.

Step 16a: For each variable, sum the NR widths of h-step ahead forecast and divide the total by NR to obtain the average width for the h-step ahead forecast for the Bonferroni and Exact procedure respectively where $h = 1, 2, 3$.

Step 17a: For each variable, count separately how many times the last three observations in the forecast period are all within the 95% simultaneous prediction intervals and divide the count by NR to obtain the coverage probability for the Bonferroni and Exact procedure respectively.

For "Linear" method:

Step 5b: Adopt the steps in section 3.3.1 and repeat the steps analogously for $y_{2,t}$. As a result, we can obtain the estimates of all the lag lengths and parameters.

Step 6b: Adopt the steps in section 3.4.3.

Step 7b: Construct the 95% simultaneous prediction intervals based on Bonferroni procedure of $y_{1,350+h}$ where $h = 1, 2, 3$. The width of the h-step ahead forecast is calculated as

$$2z_{(0.05/6)}\hat{\sigma}_{11}(h)$$

for $h = 1, 2, 3$ where $z_{(0.05/6)}$ is the $(0.05/6)$th upper percentile of the standard normal distribution.

Step 8b: Construct the 95% simultaneous prediction intervals based on the Exact procedure of $y_{1,350+h}$ where $h = 1, 2, 3$. The width of the h-step ahead forecast is calculated as

$$2c\hat{\sigma}_{11}(h)$$

for $h = 1, 2, 3$.

Step 9b: Repeat Step 6b to 8b analogously for $y_{2,t}$.

Step 10b: Repeat Step 2 to Step 4 and Step 5b to Step 9b for NR replications.

Step 11b: For each variable, sum the NR widths of h-step ahead forecast and divide the total by NR to obtain the average width for the h-step ahead forecast for the Bonferroni and Exact procedure respectively where $h = 1, 2, 3$.

Step 12b: For each variable, count separately how many times the last three observations in the forecast period are all within the 95% simultaneous prediction intervals and divide the count by NR to obtain the coverage probability for the Bonferroni and Exact procedure respectively.

Step 13b: For each variable, compute the percentage change in the average width of the h-step ahead forecast, $h = 1, 2, 3$, when using "Linear" method instead of "VAR" method for the Bonferroni and Exact procedure respectively.

Now, we outline the steps for the simulation study of Model 2.

Step 1': Input $\mathbf{v}$, $\mathbf{A}_i (i = 1, \ldots, p)$, $\boldsymbol{\Sigma}_u$ of Model 2.

Step 2': Adopt Step 2 to 4 of the above simulation.

For "VAR" method:

Step 3a': Adopt Step 5a to 10a of the above simulation.

Step 4a': Adopt the steps in section 2.4.4.

Step 5a': Adopt Step 12a to 13a of the above simulation.

Step 6a': Repeat Step 4a' to 5a' analogously for $y_{2,t}$ and $y_{3,t}$.

Step 7a': Repeat Step 2' and Step 3a' to Step 6a' for NR replications.

Step 8a': Adopt Step 16a to 17a of the above simulation.

For "Linear" method:

Step 3b': Adopt the steps in section 3.3.2 and repeat the steps analogously for $y_{2,t}$ and $y_{3,t}$.

As a result, we can obtain the estimates of all the lag lengths and parameters.

Step 4b': Adopt the steps in section 3.4.4.

Step 5b': Adopt Step 7b to 8b of the above simulation.

Step 6b': Repeat Step 4b' to 5b' analogously for $y_{2,t}$ and $y_{3,t}$.

Step 7b': Repeat Step 2' and Step 3b' to Step 6b' for NR replications.

Step 8b': Adopt Step 11b to 13b of the above simulation.

5.2 Simulation results

The procedures described in the previous section have been applied to Model 1 and Model 2 respectively. Besides, we take $NR = 10000$. The simulation results have been shown in Table 5.1a, 5.1b, 5.2a, 5.2b and 5.2c respectively.

Table 5.1a and 5.1b list the average width of the h-step ahead forecast where $h = 1, 2, 3$ and the coverage probabilities of the both methods under Bonferroni and Exact procedure respectively for both variables in Model 1. We can see that both methods perform similarly in the coverage probabilities. However, the average widths of "Linear" method are all narrower than the "VAR" method and the improvement is close to 2% in the 3-step ahead forecast for the second variable under Bonferroni and Exact procedure respectively.

Table 5.2a, 5.2b and 5.2c list the average width of the h-step ahead forecast where $h = 1, 2, 3$ and the coverage probabilities of the both methods under Bonferroni and Exact procedure respectively for all the three variables in Model 2. We can observe that both "VAR" and "Linear" methods also perform similarly in the coverage probabilities. Nevertheless, the average widths of "Linear" method are all narrower than the "VAR" method and the improvement is close to 1% on average.

In this simultaion study, it can be seen that "Linear" method is superior than the "VAR" method because it can provide shorter simultaneous prediction intervals. Also, we can observe that the average width of the Bonferroni Procedure are always wider than the corresponding width of the Exact Procedure. For comparing the both procedures, we have calculated the percentage change of the average width of the h-step ahead forecast where $h = 1, 2, 3$ when using Exact procedure instead of the Bonferroni procedure under the "Linear" method and "VAR" method respectively for both Model 1 and 2 and the results are shown in Table 5.3a, 5.3b, 5.3c and 5.3d respectively. From these tables, the coverage probabilities of two procedures are very close, but all the average widths of the Exact procedure are narrower than the widths of the Bonferroni procedure and the improvement is around 1% on average. We can conclude that the Exact procedure is preferred for both "Linear" and "VAR" method.

5.3 Concluding remarks

In this thesis, we have discussed two approaches for constructing simultaneous prediction intervals for the multple steps ahead forecasts in vector time series. One is based on the vector autoregressive model while another approach is based on a system of linear equations with exogenous variables without any restriction on the lag lengths. From the two illustrative examples and the simulation study, we have found that the approach based on a system of linear equations with exogenous variables is superior than the other approach because it can provide narrower simultaneous prediction intervals. Moreover, the Exact procedure for constructing simultaneous prediction intervals are preferred for both approaches because it can provide narrower widths, but approximately the same coverage probabilities when compared with the Bonferroni procedure.

5.4 Further research

In this thesis, there is an assumption that the white noise follows the normal distribution. For further research, we can check out the robustness of both 'VAR' and 'Linear' methods when the white noise does not follow the normal distribution assumption. For instance, we can simulate the white noise from different probability distributions such as student's t distribution and uniform distribution, and use the procedures described in the previous chapters to find out whether the proposed methods are reasonably good under these situations.

Table 5.1a Average widths and coverage probabilities for $y_{1,350+h}$, $h = 1, 2, 3$ under Model 1

Procedure	h-step-ahead period	Avg. width		% change in width	Coverage prob.	
		VAR method	Linear method		VAR method	Linear method
Bonferroni	$h = 1$	4.528	4.503	-0.552	0.943	0.944
	$h = 2$	5.070	5.029	-0.809		
	$h = 3$	5.616	5.578	-0.677		
Exact	$h = 1$	4.451	4.426	-0.562	0.937	0.939
	$h = 2$	4.983	4.943	-0.803		
	$h = 3$	5.519	5.482	-0.670		

Table 5.1b Average widths and coverage probabilities for $y_{2,350+h}$, $h = 1, 2, 3$ under Model 1

Procedure	h-step-ahead period	Avg. width		% change in width	Coverage prob.	
		VAR method	Linear method		VAR method	Linear method
Bonferroni	$h = 1$	4.267	4.252	-0.352	0.944	0.942
	$h = 2$	4.740	4.669	-1.498		
	$h = 3$	5.131	5.029	-1.988		
Exact	$h = 1$	4.223	4.213	-0.237	0.941	0.939
	$h = 2$	4.690	4.626	-1.365		
	$h = 3$	5.077	4.983	-1.851		

Table 5.2a Average widths and coverage probabilities for $y_{1,350+h}$, $h = 1, 2, 3$ under Model 2

Procedure	h-step-ahead period	Avg. width VAR method	Avg. width Linear method	% change in width	Coverage prob. VAR method	Coverage prob. Linear method
Bonferroni	$h = 1$	5.230	5.177	-1.013	0.942	0.941
	$h = 2$	5.649	5.583	-1.168		
	$h = 3$	6.116	6.052	-1.046		
Exact	$h = 1$	5.166	5.113	-1.026	0.938	0.938
	$h = 2$	5.580	5.513	-1.201		
	$h = 3$	6.041	5.976	-1.076		

Table 5.2b Average widths and coverage probabilities for $y_{2,350+h}$, $h = 1, 2, 3$ under Model 2

Procedure	h-step-ahead period	Avg. width VAR method	Avg. width Linear method	% change in width	Coverage prob. VAR method	Coverage prob. Linear method
Bonferroni	$h = 1$	4.775	4.737	-0.796	0.937	0.935
	$h = 2$	5.307	5.215	-1.734		
	$h = 3$	5.732	5.611	-2.111		
Exact	$h = 1$	4.728	4.694	-0.719	0.934	0.932
	$h = 2$	5.253	5.167	-1.637		
	$h = 3$	5.675	5.559	-2.044		

Table 5.2c Average widths and coverage probabilities for $y_{3,350+h}$, $h = 1, 2, 3$ under Model 2

Procedure	h-step-ahead period	Avg. width VAR method	Avg. width Linear method	% change in width	Coverage prob. VAR method	Coverage prob. Linear method
Bonferroni	$h = 1$	4.775	4.732	-0.901	0.944	0.942
	$h = 2$	5.357	5.303	-1.008		
	$h = 3$	5.926	5.836	-1.519		
Exact	$h = 1$	4.716	4.674	-0.891	0.940	0.938
	$h = 2$	5.291	5.238	-1.002		
	$h = 3$	5.853	5.764	-1.521		

Table 5.3a Average widths and coverage probabilities for Model 1

Using "Linear" method

	h-step-ahead period	Avg. width Bonferroni procedure	Avg. width Exact procedure	% change in width	Coverage prob. Bonferroni procedure	Coverage prob. Exact procedure
$y_{1,350+h}$	$h = 1$	4.503	4.426	-1.710	0.944	0.939
	$h = 2$	5.029	4.943	-1.710		
	$h = 3$	5.578	5.482	-1.721		
$y_{2,350+h}$	$h = 1$	4.252	4.213	-0.917	0.942	0.939
	$h = 2$	4.669	4.626	-0.921		
	$h = 3$	5.029	4.983	-0.915		

Table 5.3b Average widths and coverage probabilities for Model 2

Using "Linear" method

	h-step-ahead period	Avg. width		% change in width	Coverage prob.	
		Bonferroni procedure	Exact procedure		Bonferroni procedure	Exact procedure
$y_{1,350+h}$	$h=1$	5.177	5.113	-1.236	0.941	0.938
	$h=2$	5.583	5.513	-1.254		
	$h=3$	6.052	5.976	-1.256		
$y_{2,350+h}$	$h=1$	4.737	4.694	-0.908	0.935	0.932
	$h=2$	5.215	5.167	-0.920		
	$h=3$	5.611	5.559	-0.927		
$y_{3,350+h}$	$h=1$	4.732	4.674	-1.226	0.942	0.938
	$h=2$	5.303	5.238	-1.226		
	$h=3$	5.836	5.764	-1.234		

Table 5.3c Average widths and coverage probabilities for Model 1

Using "VAR" method

	h-step-ahead period	Avg. width		% change in width	Coverage prob.	
		Bonferroni procedure	Exact procedure		Bonferroni procedure	Exact procedure
$y_{1,350+h}$	$h=1$	4.528	4.451	-1.701	0.943	0.937
	$h=2$	5.070	4.983	-1.716		
	$h=3$	5.616	5.519	-1.727		
$y_{2,350+h}$	$h=1$	4.267	4.223	-1.031	0.944	0.941
	$h=2$	4.740	4.690	-1.055		
	$h=3$	5.131	5.077	-1.052		

Table 5.3d Average widths and coverage probabilities for Model 2

Using "VAR" method

	h-step-ahead period	Avg. width Bonferroni procedure	Avg. width Exact procedure	% change in width	Coverage prob. Bonferroni procedure	Coverage prob. Exact procedure
$y_{1,350+h}$	$h=1$	5.230	5.166	-1.224	0.942	0.938
	$h=2$	5.649	5.580	-1.221		
	$h=3$	6.116	6.041	-1.226		
$y_{2,350+h}$	$h=1$	4.775	4.728	-0.984	0.937	0.934
	$h=2$	5.307	5.253	-1.018		
	$h=3$	5.732	5.675	-0.994		
$y_{3,350+h}$	$h=1$	4.775	4.716	-1.236	0.944	0.940
	$h=2$	5.357	5.291	-1.232		
	$h=3$	5.926	5.853	-1.232		

REFERENCES

Akaike, H. (1969), Statistical Predictor Identification, *Annals of the Institute of Statistical Mathematics*, **21**, 203-217.

Akaike, H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **19**, 716-723.

Alpium, M.T. (1997), One-sided Simultaneous Prediction Intervals for AR(1) and MA(1) Processes with Exponential Innovations, *Journal of Forecasting*, **16**, 19-35.

Bowerman, B.L. and Richard, T.O'Connell (1987), *Time Series Forecasting: Unified Concepts and Computer Implementation, 2nd ed.*, Philadelphia: PWS Publishers.

Box, G.E.P. and Jenkins, G.M. (1976), *Time Series Analysis: Forecasting and Control, 2nd ed.*, Holden-Day, San Francisco, CA.

Chan, W.S. , Cheung, S.H. and Wu, K.H. (1999), Exact Joint Forecast Regions for Vector Autoregressive Models, *Journal of Applied Statistics*, **26**, 35-44.

Einhorn, H.J. and Hogarth R.M. (1982), Prediction, Diagnosis, and Causal Thinking in Forecasting, *Journal of Forecasting*, **1**, 23-26.

Genz, A. (1992), Numerical Computation of Multivariate Normal Probabilites, *Journal of Computational and Graphical Statistics*, **1**, 141-149.

Glaz, J. and Ravishanker, N. (1991), Simultaneous Prediction Intervals for Multiple Forecasts Based on Bonferroni and Product-type Inequalities, *Statistics and Probability Letters*, **12**, 57-63.

Hsiao, C. (1979), Autoregressive Modeling of Canadian Money and Income Data, *Journal of American Statistical Association*, **74**, 553-560.

Hsiao, C. (1981), Autoregressive Modeling and Money-Income Causality Detection, *Journal of Monetary Economics*, **7**, 85-106.

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis, 2nd ed.*, Germany: Springer-Verlag.

Parigi, G. and Schlitzer, G. (1995), Quarterly Forecasts of the Italian Business Cycle by Means of Monthly Economic Indicators, *Journal of Forecasting*, **14**, 117-141.

Ravishanker, N., Wu, L.S.Y. and Glaz, J. (1991), Multiplicative Prediction Intervals for Time Series: Comparsion of Simultaneous and Marginal Intervals, *Journal of Forecasting*, **10**, 445-463.

Schervish, M.J. (1984), Multivariate Normal Probabilities with Error Bound, *Applied Statistics*, **33**, 81-94.